

Igor R. Shafarevich · Alexey O. Remizov

# Linear Algebra and Geometry

# Linear Algebra and Geometry

Igor R. Shafarevich • Alexey O. Remizov

# Linear Algebra and Geometry

Translated by David Kramer and Lena Nekludova



Igor R. Shafarevich  
Steklov Mathematical Institute  
Russian Academy of Sciences  
Moscow, Russia

Alexey O. Remizov  
CMAP  
École Polytechnique CNRS  
Palaiseau Cedex, France

*Translators:*  
David Kramer  
Lancaster, PA, USA

Lena Nekludova  
Brookline, MA, USA

The original Russian edition was published as “Linejnaya algebra i geometriya” by Fizmatlit, Moscow, 2009

ISBN 978-3-642-30993-9

ISBN 978-3-642-30994-6 (eBook)

DOI 10.1007/978-3-642-30994-6

Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012946469

Mathematics Subject Classification (2010): 15-01, 51-01

© Springer-Verlag Berlin Heidelberg 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Preface

This book is the result of a series of lectures on linear algebra and the geometry of multidimensional spaces given in the 1950s through 1970s by Igor R. Shafarevich at the Faculty of Mechanics and Mathematics of Moscow State University.

Notes for some of these lectures were preserved in the faculty library, and these were used in preparing this book. We have also included some topics that were discussed in student seminars at the time. All the material included in this book is the result of joint work of both authors.

We employ in this book some results on the algebra of polynomials that are usually taught in a standard course in algebra (most of which are to be found in Chaps. 2 through 5 of this book). We have used only a few such results, without proof: the possibility of dividing one polynomial by another with remainder; the theorem that a polynomial with complex coefficients has a complex root; that every polynomial with real coefficients can be factored into a product of irreducible first- and second-degree factors; and the theorem that the number of roots of a polynomial that is not identically zero is at most the degree of the polynomial.

To provide a visual basis for this course, it was preceded by an introductory course in analytic geometry, to which we shall occasionally refer. In addition, some topics and examples are included in this book that are not really part of a course in linear algebra and geometry but are provided for illustration of various topics. Such items are marked with an asterisk and may be omitted if desired.

For the convenience of the reader, we present here the system of notation used in this book. For vector spaces we use sans serif letters:  $L, M, N, \dots$ ; for vectors, we use boldface italics:  $\mathbf{x}, \mathbf{y}, \mathbf{z}, \dots$ ; for linear transformations, we use calligraphic letters:  $\mathcal{A}, \mathcal{B}, \mathcal{C}, \dots$ ; and for the corresponding matrices, we use uppercase italic letters:  $A, B, C, \dots$ .

## Acknowledgements

The authors are grateful to M.I. Zelinkin, D.O. Orlov, and Ya.V. Tatarinov for reading parts of an earlier version of this book and making a number of useful sugges-

tions and remarks. The authors are also deeply grateful to our editor, S. Kuleshov, who gave the manuscript a very careful reading. His advice resulted in a number of important changes and additions. In particular, some parts of this book would not have appeared in their present form had it not been for his participation in this project. We would also like to offer our hearty thanks to the translators, David Kramer and Lena Nekludova, for their English translation and in particular for correcting a number of inaccuracies and typographical errors that were present in the Russian edition of this book.

# Contents

<b>1</b>	<b>Linear Equations</b>	<b>1</b>
1.1	Linear Equations and Functions	1
1.2	Gaussian Elimination	6
1.3	Examples*	15
<b>2</b>	<b>Matrices and Determinants</b>	<b>25</b>
2.1	Determinants of Orders 2 and 3	25
2.2	Determinants of Arbitrary Order	30
2.3	Properties that Characterize Determinants	37
2.4	Expansion of a Determinant Along Its Columns	39
2.5	Cramer's Rule	42
2.6	Permutations, Symmetric and Antisymmetric Functions	44
2.7	Explicit Formula for the Determinant	50
2.8	The Rank of a Matrix	53
2.9	Operations on Matrices	60
2.10	Inverse Matrices	70
<b>3</b>	<b>Vector Spaces</b>	<b>79</b>
3.1	The Definition of a Vector Space	79
3.2	Dimension and Basis	86
3.3	Linear Transformations of Vector Spaces	101
3.4	Change of Coordinates	107
3.5	Isomorphisms of Vector Spaces	112
3.6	The Rank of a Linear Transformation	118
3.7	Dual Spaces	120
3.8	Forms and Polynomials in Vectors	127
<b>4</b>	<b>Linear Transformations of a Vector Space to Itself</b>	<b>133</b>
4.1	Eigenvectors and Invariant Subspaces	133
4.2	Complex and Real Vector Spaces	142
4.3	Complexification	149
4.4	Orientation of a Real Vector Space	154

<b>5</b>	<b>Jordan Normal Form</b>	161
5.1	Principal Vectors and Cyclic Subspaces	161
5.2	Jordan Normal Form (Decomposition)	165
5.3	Jordan Normal Form (Uniqueness)	169
5.4	Real Vector Spaces	173
5.5	Applications*	176
<b>6</b>	<b>Quadratic and Bilinear Forms</b>	191
6.1	Basic Definitions	191
6.2	Reduction to Canonical Form	198
6.3	Complex, Real, and Hermitian Forms	204
<b>7</b>	<b>Euclidean Spaces</b>	213
7.1	The Definition of a Euclidean Space	213
7.2	Orthogonal Transformations	223
7.3	Orientation of a Euclidean Space*	230
7.4	Examples*	233
7.5	Symmetric Transformations	245
7.6	Applications to Mechanics and Geometry*	255
7.7	Pseudo-Euclidean Spaces	265
7.8	Lorentz Transformations	275
<b>8</b>	<b>Affine Spaces</b>	289
8.1	The Definition of an Affine Space	289
8.2	Affine Spaces	294
8.3	Affine Transformations	301
8.4	Affine Euclidean Spaces and Motions	309
<b>9</b>	<b>Projective Spaces</b>	319
9.1	Definition of a Projective Space	319
9.2	Projective Transformations	328
9.3	The Cross Ratio	335
9.4	Topological Properties of Projective Spaces*	339
<b>10</b>	<b>The Exterior Product and Exterior Algebras</b>	349
10.1	Plücker Coordinates of a Subspace	349
10.2	The Plücker Relations and the Grassmannian	353
10.3	The Exterior Product	358
10.4	Exterior Algebras*	367
10.5	Appendix*	374
<b>11</b>	<b>Quadrics</b>	385
11.1	Quadrics in Projective Space	385
11.2	Quadrics in Complex Projective Space	394
11.3	Isotropic Subspaces	398
11.4	Quadrics in a Real Projective Space	410
11.5	Quadrics in a Real Affine Space	414
11.6	Quadrics in an Affine Euclidean Space	425
11.7	Quadrics in the Real Plane*	428



<b>12</b>	<b>Hyperbolic Geometry</b>	433
12.1	Hyperbolic Space*	434
12.2	The Axioms of Plane Geometry*	443
12.3	Some Formulas of Hyperbolic Geometry*	454
<b>13</b>	<b>Groups, Rings, and Modules</b>	467
13.1	Groups and Homomorphisms	467
13.2	Decomposition of Finite Abelian Groups	475
13.3	The Uniqueness of the Decomposition	481
13.4	Finitely Generated Torsion Modules over a Euclidean Ring*	484
<b>14</b>	<b>Elements of Representation Theory</b>	497
14.1	Basic Concepts of Representation Theory	497
14.2	Representations of Finite Groups	503
14.3	Irreducible Representations	508
14.4	Representations of Abelian Groups	511
	<b>Historical Note</b>	515
	<b>References</b>	517
	<b>Index</b>	521

# Preliminaries

In this book we shall use a number of concepts from set theory. These ideas appear in most mathematics courses, and so they will be familiar to some readers. However, we shall recall them here for convenience.

## Sets and Mappings

A *set* is a collection of arbitrarily chosen objects defined by certain precisely specified properties (for example, the set of all real numbers, the set of all positive numbers, the set of solutions of a given equation, the set of points that form a given geometric figure, the set of wolves or trees in a given forest). If a set consists of a finite number of elements, then it is said to be *finite*, and if not, it is said to be *infinite*. We shall employ standard notation for certain important sets, denoting the set of natural numbers by  $\mathbb{N}$ , the set of integers by  $\mathbb{Z}$ , the set of rational numbers by  $\mathbb{Q}$ , the set of real numbers by  $\mathbb{R}$ , and the set of complex numbers by  $\mathbb{C}$ . The set of natural numbers not exceeding a given natural number  $n$ , that is, the set consisting of  $1, 2, \dots, n$ , will be denoted by  $\mathbb{N}_n$ . The objects that make up a set are called its *elements* or sometimes *points*. If  $x$  is an element of the set  $M$ , then we shall write  $x \in M$ . If we need to specify that  $x$  is not an element of  $M$ , then we shall write  $x \notin M$ .

A set  $S$  consisting of certain elements of the set  $M$  (that is, every element of the set  $S$  is also an element of the set  $M$ ) is called a *subset* of  $M$ . We write  $S \subset M$ . For example,  $\mathbb{N}_n \subset \mathbb{N}$  for arbitrary  $n$ , and likewise, we have  $\mathbb{N} \subset \mathbb{Z}$ ,  $\mathbb{Z} \subset \mathbb{Q}$ ,  $\mathbb{Q} \subset \mathbb{R}$ , and  $\mathbb{R} \subset \mathbb{C}$ . A subset of  $M$  consisting of elements  $x_\alpha \in M$  (where the index  $\alpha$  runs over a given finite or infinite set) will be denoted by  $\{x_\alpha\}$ . It is convenient to include among the subsets of a set  $M$  the set that contains no elements at all. We call this set the *empty set* and denote it by  $\emptyset$ .

Let  $M$  and  $N$  be two arbitrary sets. The collection of all elements that belong simultaneously to both  $M$  and  $N$  is called the *intersection* of  $M$  and  $N$  and is denoted by  $M \cap N$ . If we have  $M \cap N = \emptyset$ , then we say that the sets  $M$  and  $N$  are *disjoint*.

The collection of elements belonging to either  $M$  or  $N$  (or to both) is called the *union* of  $M$  and  $N$  and is denoted by  $M \cup N$ . Finally, the set of elements that belong to  $M$  but do not belong to  $N$  is called the *complement of  $N$  in  $M$*  and is denoted by  $M \setminus N$ .

We say that a set  $M$  has an *equivalence relation* defined on it if for every pair of elements  $x$  and  $y$  of  $M$ , either the elements  $x$  and  $y$  are equivalent (in which case we write  $x \sim y$ ) or they are inequivalent ( $x \not\sim y$ ), and if in addition, the following conditions are satisfied:

1. Every element of  $M$  is equivalent to itself:  $x \sim x$  (reflexivity).
2. If  $x \sim y$ , then  $y \sim x$  (symmetry).
3. If  $x \sim y$  and  $y \sim z$ , then  $x \sim z$  (transitivity).

If an equivalence relation is defined on a set  $M$ , then  $M$  can be represented as the union of a (finite or infinite) collection of sets  $M_\alpha$  called *equivalence classes* with the following properties:

- (a) Every element  $x \in M$  is contained in one and only one equivalence class  $M_\alpha$ . In other words, the sets  $M_\alpha$  are disjoint, and their union (finite or infinite) is the entire set  $M$ .
- (b) Elements  $x$  and  $y$  are equivalent ( $x \sim y$ ) if and only if they belong to the same subset  $M_\alpha$ .

Clearly, the converse holds as well: if we are given a representation of a set  $M$  as the union of subsets  $M_\alpha$  satisfying property (a), then setting  $x \sim y$  if (and only if) these elements belong to the same subset  $M_\alpha$ , we obtain an equivalence relation on  $M$ .

From the above reasoning, it is clear that the equivalence thus defined is completely abstract; there is no indication as to *precisely how* it is decided whether two elements  $x$  and  $y$  are equivalent. It is necessary only that conditions 1 through 3 above be satisfied. Therefore, on a particular set  $M$  one can define a wide variety of equivalence relations.

Let us consider a few examples. Let the set  $M$  be the natural numbers, that is,  $M = \mathbb{N}$ . Then on this set it is possible to define an equivalence relation defined by the condition that  $x \sim y$  if  $x$  and  $y$  have the same remainder on division by a given natural number  $n$ . It is clear that conditions 1 through 3 above are satisfied, and  $\mathbb{N}$  can be represented as the union of  $n$  classes (in the case  $n = 1$ , all the natural numbers are equivalent to each other and so there is only one class; if  $n = 2$ , there are two classes, namely the even numbers and the odd numbers; and so on). Now let  $M$  be the set of points in the plane or in space. We can define an equivalence relation by the rule that  $x \sim y$  if the points  $x$  and  $y$  are the same distance from a given fixed point  $O$ . Then the equivalence classes are all circles (in the case of the plane) or spheres (in space) with center at  $O$ . If, on the other hand, we wanted to consider two points equivalent if the distance between them is some given number, then we would not have an equivalence relation, since transitivity would not be satisfied.

In this book, we shall encounter several types of equivalence relations (for example, on the set of square matrices).

A *mapping* from a set  $M$  into a set  $N$  is a rule that assigns to every element of the set  $M$  a particular element of  $N$ . For example, if  $M$  is the set of all bears currently alive on Earth and  $N$  is the set of positive numbers, then assigning to each bear its weight (for example in kilograms) constitutes a mapping from  $M$  to  $N$ . We shall call such mappings of a set  $M$  into  $N$  *functions* on  $M$  with values in  $N$ . We shall usually denote such an assignment by one of the letters  $f, g, \dots$  or  $F, G, \dots$ . Mappings from a set  $M$  into a set  $N$  are indicated with an arrow and are written thus:  $f : M \rightarrow N$ . An element  $y \in N$  assigned to an element  $x \in M$  is called the *value* of the function  $f$  at the point  $x$ . This is written using an arrow with a tail,  $f : x \mapsto y$ , or the equality  $y = f(x)$ . Later on, we shall frequently display mappings between sets in the form of a *diagram*:

$$M \xrightarrow{f} N.$$

If the sets  $M$  and  $N$  coincide, then  $f : M \rightarrow M$  is called a mapping of  $M$  into *itself*. A mapping of a set into itself that assigns to each element  $x$  that same element  $x$  is called an *identity mapping*. It will be denoted by the letter  $e$ , or if it is important to specify the underlying set  $M$ , by  $e_M$ . Thus in our notation, we have  $e_M : M \rightarrow M$  and  $e_M(x) = x$  for every  $x \in M$ .

A mapping  $f : M \rightarrow N$  is called an *injection* or an *injective mapping* if different elements of the set  $M$  are assigned different elements of the set  $N$ , that is, it is injective if  $f(x_1) = f(x_2)$  always implies  $x_1 = x_2$ .

If  $S$  is a subset of  $N$  and  $f : M \rightarrow N$  is a mapping, then the collection of all elements  $x \in M$  such that  $f(x) \in S$  is called the *preimage* or *inverse image* of  $S$  and is denoted by  $f^{-1}(S)$ . In particular, if  $S$  consists of a single element  $y \in N$ , then  $f^{-1}(S)$  is called the *preimage* or *inverse image* of the element  $y$  and is written  $f^{-1}(y)$ . Using this terminology, we may say that a mapping  $f : M \rightarrow N$  is an injection if and only if for every element  $y \in N$ , its inverse image  $f^{-1}(y)$  consists of at most a single element. The words “at most” imply that certain elements  $y \in N$  may have an empty preimage. For example, let  $M = N = \mathbb{R}$  and suppose the mapping  $f$  assigns to each real number  $x$  the value  $f(x) = \arctan x$ . Then  $f$  is injective, since the inverse image  $f^{-1}(y)$  consists of a single element if  $|y| < \frac{\pi}{2}$  and is the empty set if  $|y| \geq \frac{\pi}{2}$ .

If  $S$  is a subset of  $M$  and  $f : M \rightarrow N$  is a mapping, then the collection of all elements  $y \in N$  such that  $y = f(x)$  for some  $x \in S$  is called the *image* of the subset  $S$  and is denoted by  $f(S)$ . In particular, the subset  $S$  could be the entire set  $M$ , in which case  $f(M)$  is called the *image* of the mapping  $f$ . We note that the image of  $f$  does not have to consist of the entire set  $N$ . For example, if  $M = N = \mathbb{R}$  and  $f$  is the squaring operation (raising to the second power), then  $f(M)$  is the set of nonnegative real numbers and does not coincide with the set  $\mathbb{R}$ .

If again  $S$  is a subset of  $M$  and  $f : M \rightarrow N$  a mapping, then applying the mapping only to elements of the set  $S$  defines a mapping  $f : S \rightarrow N$ , called the *restriction* of the mapping  $f$  to  $S$ . In other words, the restriction mapping is defined by taking  $f(x)$  for each  $x \in S$  as before and simply ignoring all  $x \notin S$ . Conversely, if we start off with a mapping  $f : S \rightarrow N$  defined only on the subset  $S$ , and then somehow define  $f(x)$  for the remaining elements  $x \in M \setminus S$ , then we obtain a mapping  $f : M \rightarrow N$ , called an *extension* of  $f$  to  $M$ .

A mapping  $f : M \rightarrow N$  is *bijective* or a *bijection* if it is injective and the image  $f(M)$  is the entire set  $N$ , that is,  $f(M) = N$ . Equivalently, a mapping is a bijection if for each element  $y \in N$ , there exists precisely one element  $x \in M$  such that  $y = f(x)$ .<sup>1</sup> In this case, it is possible to define a mapping from  $N$  into  $M$  that assigns to each element  $y \in N$  the unique element  $x \in M$  such that  $f(x) = y$ . Such a mapping is called the *inverse* of  $f$  and is denoted by  $f^{-1} : N \rightarrow M$ . Now suppose we are given sets  $M, N, L$  and mappings  $f : M \rightarrow N$  and  $g : N \rightarrow L$ , which we display in the following diagram:

$$M \xrightarrow{f} N \xrightarrow{g} L. \quad (1)$$

Then application of  $f$  followed by  $g$  defines a mapping from  $M$  to  $L$  by the obvious rule: first apply the mapping  $f : M \rightarrow N$ , which assigns to each element  $x \in M$  an element  $y \in N$ , and then apply the mapping  $g : N \rightarrow L$  that takes an element  $y$  to some element  $z \in L$ . We thus obtain a mapping from  $M$  to  $L$  called the *composition* of the mappings  $f$  and  $g$ , written  $g \circ f$  or simply  $gf$ . Using this notation, the composition mapping is defined by the formula

$$(g \circ f)(x) = g(f(x)) \quad (2)$$

for an arbitrary  $x \in M$ . We note that in equation (2), the letters  $f$  and  $g$  that denote the two mappings appear in the reverse order to that in the diagram (1). As we shall see later, such an arrangement has a number of advantages.

As an example of the composition of mappings we offer the obvious equalities

$$e_N \circ f = f, \quad f \circ e_M = f,$$

valid for any mapping  $f : M \rightarrow N$ , and likewise the equalities

$$f \circ f^{-1} = e_N, \quad f^{-1} \circ f = e_M,$$

which are valid for any bijective mapping  $f : M \rightarrow N$ .

The composition of mappings has an important property. Suppose that in addition to the mapping shown in diagram (1), we have as well a mapping  $h : L \rightarrow K$ , where  $K$  is an arbitrary set. Then we have

$$h \circ (g \circ f) = (h \circ g) \circ f. \quad (3)$$

The truth of this claim follows at once from the definitions. First of all, it is apparent that both sides of equation (3) contain a mapping from  $M$  to  $K$ . Thus we need to show that when applied to any element  $x \in M$ , both sides give the same element of the set  $K$ . According to definition (2), for the left-hand side of (3), we obtain

$$h \circ (g \circ f)(x) = h((g \circ f)(x)), \quad (g \circ f)(x) = g(f(x)).$$

---

<sup>1</sup>*Translator's note:* The term *one-to-one* is also used in this context. However, its use can be confusing: an injection is sometimes called a *one-to-one mapping*, while a bijection is sometimes called a *one-to-one correspondence*. In this book, we shall strive to stick to the terms *injective* and *bijective*.

Substituting the second equation into the first, we finally obtain  $h \circ (g \circ f)(x) = h(g(f(x)))$ . Analogous reasoning shows that we obtain precisely the same expression for the right-hand side of equation (3).

The property expressed by formula (3) is called *associativity*. Associativity plays an important role, both in this course and in other branches of mathematics. Therefore, we shall pause here to consider this concept in more detail. For the sake of generality, we shall consider a set  $M$  of arbitrary objects (they can be numbers, matrices, mappings, and so on) on which is defined the operation of multiplication associating two elements  $a \in M$  and  $b \in M$  with some element  $ab \in M$ , which we call the *product*, such that it possesses the associative property:

$$(ab)c = a(bc). \quad (4)$$

The point of condition (4) is that without it, we can calculate the product of elements  $a_1, \dots, a_m$  for  $m > 2$  only if the sequence of multiplications is indicated by parentheses, indicating which pairs of adjacent elements we are allowed to multiply. For example, with  $m = 3$ , we have two possible arrangements of the parentheses:  $(a_1 a_2) a_3$  and  $a_1 (a_2 a_3)$ . For  $m = 4$  we have five variants:

$$\begin{aligned} & ((a_1 a_2) a_3) a_4, & (a_1 (a_2 a_3)) a_4, & (a_1 a_2) (a_3 a_4), \\ & a_1 ((a_2 a_3) a_4), & a_1 (a_2 (a_3 a_4)), \end{aligned}$$

and so on. It turns out that if for three factors ( $m = 3$ ), the product does not depend on how the parentheses are ordered (that is, the associative property is satisfied), then it will be independent of the arrangement of parentheses with any number of factors.

This assertion is easily proved by induction on  $m$ . Indeed, let us suppose that it is true for all products of  $m$  or fewer elements, and let us consider products of  $m + 1$  elements  $a_1, \dots, a_m, a_{m+1}$  for all possible arrangements of parentheses. It is easily seen that in this case, there are two possible alternatives: either there is no parenthesis between elements  $a_m$  and  $a_{m+1}$ , or else there is one. Since by the induction hypothesis, the assertion is correct for  $a_1, \dots, a_m$ , then in the first case we obtain the product  $(a_1 \cdots a_{m-1})(a_m a_{m+1})$ , while in the second case, we have  $(a_1 \cdots a_m) a_{m+1} = ((a_1 \cdots a_{m-1}) a_m) a_{m+1}$ . Introducing the notation  $a = a_1 \cdots a_{m-1}$ ,  $b = a_m$ , and  $c = a_{m+1}$ , we obtain the products  $a(bc)$  and  $(ab)c$ , the equality of which follows from property (4).

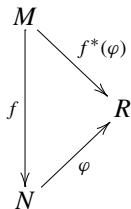
In the special case  $a_1 = \cdots = a_m = a$ , the product  $a_1 \cdots a_m$  is denoted by  $a^m$  and is called the  $m$ th power of the element  $a$ .

There is another important concept connected to the composition of mappings.

Let  $R$  be a given set. We shall denote by  $\mathfrak{F}(M, R)$  the collection of all mappings  $M \rightarrow R$ , and analogously, by  $\mathfrak{F}(N, R)$  the collection of all mappings  $N \rightarrow R$ . Then with every mapping  $f : M \rightarrow N$  is associated the particular mapping  $f^* : \mathfrak{F}(N, R) \rightarrow \mathfrak{F}(M, R)$ , called the *dual* to  $f$  and defined as follows: For every mapping  $\varphi \in \mathfrak{F}(N, R)$  it assigns the mapping  $f^*(\varphi) \in \mathfrak{F}(M, R)$  according to the formula

$$f^*(\varphi) = \varphi \circ f. \quad (5)$$

Formula (5) indicates that for an arbitrary element  $x \in M$ , we have the equality  $f^*(\varphi)(x) = \varphi \circ f(x)$ , which can also be expressed by the following diagram:



Here we become acquainted with the following general mathematical fact: *Functions are written in reverse order in comparison with the order of the sets on which they are defined.* This phenomenon will appear in our book, as well as in other courses in relationship to more complex objects (such as differential forms).

The dual mapping  $f^*$  possesses the following important property: If we have mappings of sets, as depicted in diagram (1), then

$$(g \circ f)^* = f^* \circ g^*. \quad (6)$$

Indeed, we obtain the dual mappings

$$\mathfrak{F}(L, R) \xrightarrow{g^*} \mathfrak{F}(N, R) \xrightarrow{f^*} \mathfrak{F}(M, R).$$

By definition, for  $g \circ f : M \rightarrow L$ , the dual mapping  $(g \circ f)^*$  is a mapping from  $g \circ \mathfrak{F}(L, R)$  into  $\mathfrak{F}(M, R)$ . As can be seen from (2),  $f^* \circ g^*$  is also a mapping of the same sets. It remains for us to show that  $(g \circ f)^*$  and  $f^* \circ g^*$  take every element  $\psi \in \mathfrak{F}(L, R)$  to one and the same element of the set  $\mathfrak{F}(M, R)$ . By (5), we have

$$(g \circ f)^*(\psi) = \psi \circ (g \circ f).$$

Analogously, taking into account (2), we obtain the relationship

$$f^* \circ g^*(\psi) = f^*(g^*(\psi)) = f^*(\psi \circ g) = (\psi \circ g) \circ f.$$

Thus for a proof of equality (6), it suffices to verify associativity:  $\psi \circ (g \circ f) = (\psi \circ g) \circ f$ .

Up to now, we have considered mappings (functions) of a single argument. The definition of functions of several arguments is reduced to this notion with the help of the operation of *product* of sets.

Let  $M_1, \dots, M_n$  be arbitrary sets. Consider the *ordered* collection  $(x_1, \dots, x_n)$ , where  $x_i$  is an arbitrary element of the set  $M_i$ . The word “ordered” indicates that in such collections, the order of the sequence of elements  $x_i$  is taken into account. For example, in the case  $n = 2$  and  $M_1 = M_2$ , the pairs  $(x_1, x_2)$  and  $(x_2, x_1)$  are considered to be different if  $x_1 \neq x_2$ . A set consisting of all ordered collections  $(x_1, \dots, x_n)$  is called the *product* of the sets  $M_1, \dots, M_n$  and is denoted by  $M_1 \times \dots \times M_n$ .

In the special case  $M_1 = \dots = M_n = M$ , the product  $M_1 \times \dots \times M_n$  is denoted by  $M^n$  and is called the  *$n$ th power* of the set  $M$ .

Now we can define a function of an arbitrary number of arguments, each of which assumes values from “its own” set. Let  $M_1, \dots, M_n$  be arbitrary sets, and let us

define  $M = M_1 \times \cdots \times M_n$ . By definition, the mapping  $f : M \rightarrow N$  assigns to each element  $x \in M$  a certain element  $y \in N$ , that is, it assigns to  $n$  elements  $x_1 \in M_1, \dots, x_n \in M_n$ , taken in the assigned order, the element  $y = f(x_1, \dots, x_n)$  of the set  $N$ . This is a function of  $n$  arguments  $x_i$ , each of which takes values from “its own” set  $M_i$ .

## Some Topological Notions

Up to now, we have been speaking about sets of arbitrary form, not assuming that they possess any additional properties. Generally, that will not suffice. For example, let us assume that we wish to compare two geometric figures, in particular, to determine the extent to which they are or are not “alike.” Let us consider the two figures to be sets whose elements are points in a plane or in space. If we wish to limit ourselves to the concepts introduced above, then it is natural to consider “alike” those sets between which there exists a bijection. However, toward the end of the nineteenth century, Georg Cantor demonstrated that there exists a bijection between the points of a line segment and those of the interior of a square.<sup>2</sup> At the same time, Richard Dedekind conjectured that our intuitive idea of “alike” of figures is connected with the possibility of establishing between them a *continuous* bijection. But for that, it is necessary to define what it means for a mapping to be continuous.

The branch of mathematics in which one studies continuous mappings of abstract sets and considers objects with a precision only up to bijective continuous mappings is called *topology*. Using the words of Hermann Weyl, we may say that in this book, “the mountain range of topology will loom on the horizon.” More precisely, we shall introduce some topological notions only now and then, and then only the simplest ones. We shall formulate them now, but we shall appeal to them seldom, and only to indicate a connection between the objects that we are considering with other branches of mathematics to which the reader may be introduced in more detail in other courses or textbooks. Such instances can be read or passed over as desired; they will not be used in the remainder of the book. To define a continuous mapping  $f : M \rightarrow N$  it is necessary first to define the notion of *convergence* on the sets  $M$  and  $N$ . In some cases, we will define convergence on sets (for example, in spaces of vectors, spaces of matrices, or projective spaces), based on the notion of convergence in  $\mathbb{R}$  and  $\mathbb{C}$ , which is assumed to be familiar to the reader from a course in calculus. In other cases, we shall make use of the notion of *metric*.

A set  $M$  is called a *metric space* if there exists a function  $r : M^2 \rightarrow \mathbb{R}$  assigning to every pair of points  $x, y \in M$  a number  $r(x, y)$  that satisfies the following conditions:

1.  $r(x, y) > 0$  for  $x \neq y$ , and  $r(x, x) = 0$ , for every  $x, y \in M$ .

---

<sup>2</sup>This result so surprised him, that as Cantor wrote in a letter, he believed for a long time that it was incorrect.



2.  $r(x, y) = r(y, x)$  for every  $x, y \in M$ .
3. For any three points  $x, y, z \in M$  one has the inequality

$$r(x, z) \leq r(x, y) + r(y, z). \quad (7)$$

Such a function  $r(x, y)$  is called a *metric* or *distance* on  $M$ , and the properties enumerated in its definition constitute an axiomatization of the usual properties of distance known from courses in elementary or analytic geometry.

For example, the set  $\mathbb{R}$  of all real numbers (and also any subset of it) becomes a metric space if for every pair of numbers  $x$  and  $y$  we introduce the function  $r(x, y) = |x - y|$  or  $r(x, y) = \sqrt{|x - y|}$ .

For an arbitrary metric space there is automatically defined the notion of *convergence* of points in the space: a sequence of points  $x_k$  *converges to the point*  $x$  as  $k \rightarrow \infty$  (notation:  $x_k \rightarrow x$ ) if  $r(x_k, x) \rightarrow 0$  as  $k \rightarrow \infty$ . The point  $x$  in this case is called the *limit* of the sequence  $x_k$ .

Let  $X \subset M$  be some subset of  $M$ , and  $M$  a metric space with the metric  $r(x, y)$ , that is, a mapping  $r : M^2 \rightarrow \mathbb{R}$  satisfying the three properties given above. It is clear that the restriction of  $r(x, y)$  to the subset  $X^2 \subset M^2$  also satisfies those properties, and hence it defines a metric on  $X$ . We say that  $X$  is a metric space with the metric *induced* by the metric of the enclosing space  $M$  or that  $X \subset M$  is a metric *subspace*.

The subset  $X$  is said to be *closed* in  $M$  if it contains the limit point of every convergent sequence in  $X$ , and it is said to be *bounded* if there exist a point  $x \in X$  and a number  $c > 0$  such that  $r(x, y) \leq c$  for all  $y \in X$ .

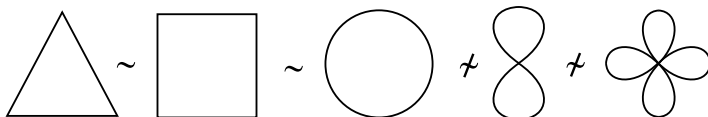
Let  $M$  and  $N$  be sets on each of which is defined the notion of convergence (for example,  $M$  and  $N$  could be metric spaces). A mapping  $f : M \rightarrow N$  is said to be *continuous* at the point  $x \in M$  if for every convergent sequence  $x_k \rightarrow x$  of points in the set  $M$ , one has  $f(x_k) \rightarrow f(x)$ . If the mapping  $f : M \rightarrow N$  is continuous at every point  $x \in M$ , then we say that it is *continuous on the set*  $M$  or simply that it is *continuous*.

The mapping  $f : M \rightarrow N$  is called a *homeomorphism* if it is injective with an injective inverse mapping  $f^{-1} : N \rightarrow M$ , both of which are continuous.<sup>3</sup> The sets  $M$  and  $N$  are said to be *homeomorphic* or *topologically equivalent* if there exists a homeomorphism  $f : M \rightarrow N$ . It is easily seen that the property among sets of being homeomorphic (for a given fixed definition of convergence) is an equivalence relation.

Given two infinite sets  $M$  and  $N$  on which no metrics have initially been defined, if we then supply them with metrics using first one definition and then another, we will obtain differing notions of homeomorphism  $f : M \rightarrow N$ , and it can turn out that in one type of metric,  $M$  and  $N$  are homeomorphic, while in another type they are not. For example, on arbitrary sets  $M$  and  $N$  let us define what is called the *discrete* metric, defined by the relations  $r(x, y) = 1$  for all  $x \neq y$  and  $r(x, x) = 0$  for all  $x$ . It is clear that with such a definition, all the properties of a metric are

---

<sup>3</sup>We wish to emphasize that this last condition is essential: from the continuity of  $f$  one may not conclude the continuity of  $f^{-1}$ .



**Fig. 1** Homeomorphic and nonhomeomorphic curves (the symbol  $\sim$  means that the figures are homeomorphic, while  $\neq$  means that they are not)

satisfied, but the notion of homeomorphism  $f : M \rightarrow N$  becomes empty: it simply coincides with the notion of bijection. For indeed, in the discrete metric, a sequence  $x_k$  converges to  $x$  if beginning with some index  $k$ , all the points  $x_k$  are equal to  $x$ . As follows from the definition of continuous mapping given above, this means that every mapping  $f : M \rightarrow N$  is continuous.

For example, according to a theorem of Cantor, a line segment and a square are homeomorphic under the discrete metric, but if we consider them, for example, as metric spaces in the plane on which distance is defined as in a course in elementary geometry (let us say using the system of Cartesian coordinates), then the two sets are no longer homeomorphic.

This shows that the discrete metric fails to reflect some important properties of distance with which we are familiar from courses in geometry, one of which is that for an arbitrarily small number  $\varepsilon > 0$ , there exist two distinct points  $x$  and  $y$  for which  $r(x, y) < \varepsilon$ . Therefore, if we are to formulate our intuitive idea of “geometric similarity” of two sets  $M$  and  $N$ , it is necessary to consider them not with an arbitrary metric, but with a metric that reflects these geometric notions.

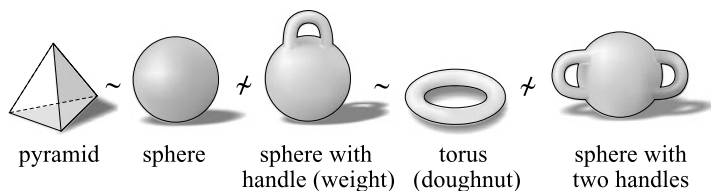
We are not going to go more deeply into this question, since for our purposes that is unnecessary. In this book, when we “compare” sets  $M$  and  $N$ , where at least one of them (say  $N$ ) is a geometric figure in the plane (or in space), then distance will be determined in the usual way, with the metric on  $N$  induced by the metric in the plane (or in the space) in which it lies. It remains for us to define the metric (or notion of convergence) on the set  $M$  in such a way that  $M$  and  $N$  are homeomorphic. That is how we shall make precise the idea of comparison.

If the figures  $M$  and  $N$  are metric subspaces of the plane or space with distance defined as in elementary geometry, then there exists for them a very graphic interpretation of the concept of topological equivalence. Imagine that figures  $M$  and  $N$  are made out of rubber. Then their being homeomorphic means that we can deform  $M$  into  $N$  without tearing and without gluing together any points. This last condition (“without tearing and without gluing together any points”) is what makes the notion of homeomorphism much stronger than simply a bijective mapping of sets.

For example, an arbitrary continuous closed curve without self-intersection (for example, a triangle or square) is homeomorphic to a circle. On the other hand, a continuous closed curve with self-intersection (say a figure eight) is not homeomorphic to a circle (see Fig. 1).

In Fig. 2 we have likewise depicted examples of homeomorphic and nonhomeomorphic figures, this time in three-dimensional space.

We conclude by introducing a few additional simple topological concepts that will be used in this book.



**Fig. 2** Homeomorphic and nonhomeomorphic surfaces

A *path* in a metric space  $M$  is a continuous mapping  $f : I \rightarrow M$ , where  $I$  is an interval of the real line. Without any loss of generality, we may assume that  $I = [0, 1]$ . In this case, the points  $f(0)$  and  $f(1)$  are called the *beginning* and *end* of the path. Two points  $x, y \in M$  are said to be *continuously deformable* into each other if there is a path in which  $x$  is the beginning and  $y$  is the end. Such a path is called a *deformation* of  $x$  into  $y$ , and we shall notate the fact that  $x$  and  $y$  are deformable into one another by  $x \sim y$ .

The property for elements of a space  $M$  to be continuously deformable into one another is an equivalence relation on  $M$ , since properties 1 through 3 that define such a relation are satisfied. Indeed, the reflexive property is obvious. To prove symmetry, it suffices to observe that if  $f(t)$  is a deformation of  $x$  into  $y$ , then  $f(1 - t)$  is a deformation of  $y$  into  $x$ . Now let us verify transitivity. Let  $x \sim y$  and  $y \sim z$ ,  $f(t)$  a deformation of  $x$  into  $y$ , and  $g(t)$  a deformation of  $y$  into  $z$ . Then the mapping  $h : I \rightarrow M$  determined by the equality  $h(t) = f(2t)$  for  $t \in [0, \frac{1}{2}]$  and the equality  $h(t) = g(2t - 1)$  for  $t \in [\frac{1}{2}, 1]$  is continuous, and for this mapping, the equalities  $h(0) = f(0) = x$ ,  $h(1) = g(1) = z$  are satisfied. Thus  $h(t)$  gives the continuous deformation of the point  $x$  to  $z$ , and therefore we have  $x \sim z$ .

If every pair of elements of a metric space  $M$  can be deformed one into the other (that is, the relationship  $\sim$  defines a single equivalence class), then the space  $M$  is said to be *path-connected*. If that is not the case, then for each element  $x \in M$  we consider the equivalence class  $M_x$  consisting of all elements  $y \in M$  such that  $x \sim y$ . By the definition of equivalence class, the metric space  $M_x$  will be path-connected. It is called the *path-connected component* of the space  $M$  containing the point  $x$ . Thus the equivalence relation defined by a continuous deformation decomposes  $M$  into path-connected components.

In a number of important cases, the number of components is finite, and we obtain the representation  $M = M_1 \cup \dots \cup M_k$ , where  $M_i \cap M_j = \emptyset$  for  $i \neq j$  and each  $M_i$  is path-connected. It is easily seen that such a representation is unique. The sets  $M_i$  are called the *path-connected components* of the space  $M$ .

For example, a hyperboloid of one sheet, a sphere, and a cone are each path-connected, but a hyperboloid of two sheets is not: it has two path-connected components. The set of real numbers defined by the condition  $0 < |x| < 1$  has two path-connected components (one containing positive numbers; the other, negative numbers), while the set of complex numbers defined by the same condition is path-connected. The properties preserved by homeomorphisms are called *topological*

properties. Thus, for example, the property of path-connectedness is topological, as is the number of path-connected components.

Let  $M$  and  $N$  be metric spaces (let us denote their respective metrics by  $r$  and  $r'$ ). A mapping  $f : M \rightarrow N$  is called an *isometry* if it is bijective and preserves distances between points, that is,

$$r(x_1, x_2) = r'(f(x_1), f(x_2)) \quad (8)$$

for every pair of points  $x_1, x_2 \in M$ . From the relationship (8), it follows automatically that an isometry is an embedding. Indeed, if there existed points  $x_1 \neq x_2$  in the set  $M$  for which the equation  $f(x_1) = f(x_2)$  were satisfied, then from condition 1 in the definition of a metric space, the left-hand side of (8) would be different from zero, while the right-hand side would be equal to zero. Therefore, the requirement of a bijective mapping is here reduced to the condition that the image of  $f(M)$  coincide with all of the set  $N$ .

Metric spaces  $M$  and  $N$  are called *isometric* or *metrically equivalent* if there exists an isometry  $f : M \rightarrow N$ . It is easy to see that an isometry is a homeomorphism and generalizes the notion of the motion of a rigid body in space, whereby we cannot arbitrarily deform the sets  $M$  and  $N$  into one another as if they were made of rubber (without tearing and gluing). We can only treat them as if they were rigid or made of flexible, but not compressible or stretchable, materials (for example, an isometry of a piece of paper is obtained by bending it or rolling it up).

In the plane or in space with distance determined by the familiar methods of elementary geometry, examples of isometries are parallel translations, rotations, and symmetry transformations. Thus, for example, two triangles in the plane are isometric if and only if they are “equal” (that is, congruent in the sense defined in courses in school geometry, namely equality of sides and angles), and two ellipses are isometric if and only if they have equal major and minor axes.

In conclusion, we observe that in the definition of homeomorphism, path-connectedness, and path-connected component, the notion of metric played only an auxiliary role. We used it to define the notion of *convergence* of a sequence of points, so that we could speak of continuity of a mapping and thereby introduce concepts that depend on this notion. It is convergence that is the basic topological notion. It can be defined by various metrics, and it can also be defined in another way, as is usually done in topology.

# Chapter 1

## Linear Equations

## 1.1 Linear Equations and Functions

In this chapter, we will be studying systems of equations of degree one. We shall let the number of equations and number of unknowns be arbitrary. We begin by choosing suitable notation. Since the number of unknowns can be arbitrarily large, it will not suffice to use the twenty-six letters of the alphabet:  $x, y, \dots, z$ , and so on. Therefore, we shall use a single letter to designate all the unknowns and distinguish among them with an index, or subscript:  $x_1, x_2, \dots, x_n$ , where  $n$  is the number of unknowns. The coefficients of our equations will be notated using the same principle, and a single equation of the first degree will be written thus:

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b. \quad (1.1)$$

A first-degree equation is also called a *linear* equation.

We shall use the same principle to distinguish among the various equations. But since we have already used one index for designating the coefficients of the unknowns, we introduce a second index. We shall denote the coefficient of  $x_k$  in the  $i$ th equation by  $a_{ik}$ . To the right side of the  $i$ th equation we attach the symbol  $b_i$ . Therefore, the  $i$ th equation is written

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i, \quad (1.2)$$

and a system of  $m$  equations in  $n$  unknowns will look like this:

[illegible]

The numbers  $b_1, \dots, b_m$  are called the *constant terms* or just *constants* of the system (1.3). It will sometimes be convenient to focus our attention on the coefficients of

the unknowns in system (1.3), and then we shall use the following tableau:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad (1.4)$$

with  $m$  rows and  $n$  columns. Such a rectangular array of numbers is called an  $m \times n$  *matrix* or a *matrix of type*  $(m, n)$ , and the numbers  $a_{ij}$  are called the *elements* of the matrix. If  $m = n$ , then the matrix is an  $n \times n$  *square matrix*. In this case, the elements  $a_{11}, a_{22}, \dots, a_{nn}$ , each located in a row and column with the same index, form the matrix's *main diagonal*.

The matrix (1.4), whose elements are the coefficients of the unknowns of system (1.3), is called the *matrix associated with the system*. Along with the matrix (1.4), it is frequently necessary to consider the matrix that includes the constant terms:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}. \quad (1.5)$$

This matrix has one column more than matrix (1.4), and thus it is an  $m \times (n + 1)$  matrix. Matrix (1.5) is called the *augmented matrix of the system* (1.3).

Let us consider in greater detail the left-hand side of equation (1.1). Here we are usually talking about trying to find specific values of the unknowns  $x_1, \dots, x_n$  that satisfy the relationship (1.1). But it is also possible to consider the expression  $a_1x_1 + a_2x_2 + \cdots + a_nx_n$  from another point of view. We can substitute arbitrary numbers

$$x_1 = c_1, \quad x_2 = c_2, \quad \dots, \quad x_n = c_n, \quad (1.6)$$

for the unknowns  $x_1, x_2, \dots, x_n$  in the expression, each time obtaining as a result a certain number

$$a_1c_1 + a_2c_2 + \cdots + a_nc_n. \quad (1.7)$$

From this point of view, we are dealing with a certain type of function. In the given situation, the initial element to which we are associating something is the set of values (1.6), which is determined simply by the set of numbers  $(c_1, c_2, \dots, c_n)$ . We shall call such a set of numbers a *row* of length  $n$ . It is the same as a  $1 \times n$  matrix. We associate the expression (1.7), which is a number, with the row  $(c_1, c_2, \dots, c_n)$ . Then employing the notation of page xiii, we obtain a function on the set  $M$  with values in  $N$ , where  $M$  is the set of all rows of length  $n$ , and  $N$  is the set of all numbers.

**Definition 1.1** A function  $F$  on the set of all rows of length  $n$  with values in the set of all numbers is said to be *linear* if there exist numbers  $a_1, a_2, \dots, a_n$  such that  $F$  associates to each row  $(c_1, c_2, \dots, c_n)$  the number (1.7).

We shall proceed to denote a row by a single boldface italic letter, such as  $\mathbf{c}$ , and shall associate with it a number,  $F(\mathbf{c})$ , via the linear function  $F$ . Thus if  $\mathbf{c} = (c_1, c_2, \dots, c_n)$ , then  $F(\mathbf{c}) = a_1c_1 + a_2c_2 + \dots + a_nc_n$ .

In the case  $n = 1$ , a linear function coincides with the well-known concept of direct proportionality, which will be familiar to the reader from secondary-school mathematics. Thus the notion of linear function is a natural generalization of direct proportionality. To emphasize this analogy, we shall define some operations on rows of length  $n$  in analogy to arithmetic operations on numbers.

**Definition 1.2** Let  $\mathbf{c}$  and  $\mathbf{d}$  be rows of a fixed length  $n$ , that is,

$$\mathbf{c} = (c_1, c_2, \dots, c_n), \quad \mathbf{d} = (d_1, d_2, \dots, d_n).$$

Their *sum* is the row  $(c_1 + d_1, c_2 + d_2, \dots, c_n + d_n)$ , denoted by  $\mathbf{c} + \mathbf{d}$ . The *product* of row  $\mathbf{c}$  and the number  $p$  is the row  $(pc_1, pc_2, \dots, pc_n)$ , denoted by  $p\mathbf{c}$ .

**Theorem 1.3** A function  $F$  on the set of rows of length  $n$  is linear if and only if it possesses the following properties:

$$F(\mathbf{c} + \mathbf{d}) = F(\mathbf{c}) + F(\mathbf{d}), \tag{1.8}$$

$$F(p\mathbf{c}) = pF(\mathbf{c}), \tag{1.9}$$

for all rows  $\mathbf{c}, \mathbf{d}$  and all numbers  $p$ .

*Proof* Properties (1.8) and (1.9) are the direct analogue of the well-known conditions for direct proportionality.

The proof of properties (1.8) and (1.9) is completely obvious. Let the linear function  $F$  associate to each row  $\mathbf{c} = (c_1, c_2, \dots, c_n)$  the number (1.7). By the above definition, the sum of rows  $\mathbf{c} = (c_1, \dots, c_n)$  and  $\mathbf{d} = (d_1, \dots, d_n)$  is the row  $\mathbf{c} + \mathbf{d} = (c_1 + d_1, \dots, c_n + d_n)$ , and it follows that

$$\begin{aligned} F(\mathbf{c} + \mathbf{d}) &= a_1(c_1 + d_1) + \dots + a_n(c_n + d_n) \\ &= (a_1c_1 + a_1d_1) + \dots + (a_nc_n + a_nd_n) \\ &= (a_1c_1 + \dots + a_nc_n) + (a_1d_1 + \dots + a_nd_n) \\ &= F(\mathbf{c}) + F(\mathbf{d}), \end{aligned}$$

which is equation (1.8). In exactly the same way, we obtain

$$F(p\mathbf{c}) = a_1(pc_1) + \dots + a_n(pc_n) = p(a_1c_1 + \dots + a_nc_n) = pF(\mathbf{c}).$$

Let us now prove the reverse assertion: any function  $F$  on the set of rows of length  $n$  with numerical values satisfying properties (1.8) and (1.9) is linear. To show this, let us consider the row  $\mathbf{e}_i$  in which every entry except the  $i$ th is equal to zero, while the  $i$ th is equal to 1, that is,  $\mathbf{e}_i = (0, \dots, 1, \dots, 0)$ , where the 1 is in the  $i$ th place.

Let us set  $F(\mathbf{e}_i) = a_i$  and let us prove that for an arbitrary row  $\mathbf{c} = (c_1, \dots, c_n)$ , the following equality is satisfied:  $F(\mathbf{c}) = a_1c_1 + \dots + a_nc_n$ . From that we will be able to conclude that the function  $F$  is linear.

For this, let us convince ourselves that  $\mathbf{c} = c_1\mathbf{e}_1 + \dots + c_n\mathbf{e}_n$ . This is almost obvious: let us consider what number is located at the  $i$ th place in the row  $c_1\mathbf{e}_1 + \dots + c_n\mathbf{e}_n$ . In any row  $\mathbf{e}_k$  with  $k \neq i$ , there is a 0 in the  $i$ th place, and therefore, the same is true for  $c_k\mathbf{e}_k$ , which means that in the row  $c_i\mathbf{e}_i$ , the element  $c_i$  is located at the  $i$ th place. As a result, in the complete sum  $c_1\mathbf{e}_1 + \dots + c_n\mathbf{e}_n$ , there is  $c_i$  at the  $i$ th place. This is true for arbitrary  $i$ , which implies that the sum under consideration coincides with the row  $\mathbf{c}$ .

Now let us consider  $F(\mathbf{c})$ . Using properties (1.8) and (1.9)  $n$  times, we obtain

$$\begin{aligned} F(\mathbf{c}) &= F(c_1\mathbf{e}_1) + F(c_2\mathbf{e}_2 + \dots + c_n\mathbf{e}_n) = c_1F(\mathbf{e}_1) + F(c_2\mathbf{e}_2 + \dots + c_n\mathbf{e}_n) \\ &= a_1c_1 + F(c_2\mathbf{e}_2 + \dots + c_n\mathbf{e}_n) = a_1c_1 + a_2c_2 + F(c_3\mathbf{e}_3 + \dots + c_n\mathbf{e}_n) \\ &= \dots = a_1c_1 + a_2c_2 + \dots + a_nc_n, \end{aligned}$$

as asserted. □

We shall soon convince ourselves of the usefulness of these properties of a linear function. Let us define the operations on linear functions that we shall be meeting in the sequel.

**Definition 1.4** Let  $F$  and  $G$  be two linear functions on the set of rows of length  $N$ . Their *sum* is the function  $F + G$ , on the same set, defined by the equality  $(F + G)(\mathbf{c}) = F(\mathbf{c}) + G(\mathbf{c})$  for every row  $\mathbf{c}$ . The *product* of the linear function  $F$  and the number  $p$  is the function  $pF$ , defined by the relation  $(pF)(\mathbf{c}) = p \cdot F(\mathbf{c})$ .

Using Theorem 1.3, we obtain that both  $F + G$  and  $pF$  are linear functions.

We return now to the system of linear equations (1.3). Clearly, it can be written in the form

$$\begin{cases} F_1(\mathbf{x}) = b_1, \\ \dots \\ F_m(\mathbf{x}) = b_m, \end{cases} \quad (1.10)$$

where  $F_1(\mathbf{x}), \dots, F_m(\mathbf{x})$  are linear functions defined by the relationships

$$F_i(\mathbf{x}) = a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n.$$

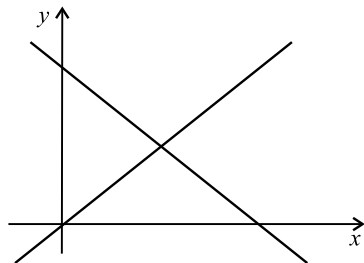
A row  $\mathbf{c}$  is called a *solution* of the system (1.10) if on substituting  $\mathbf{x}$  by  $\mathbf{c}$ , all the equations are transformed into identities, that is,  $F_1(\mathbf{c}) = b_1, \dots, F_m(\mathbf{c}) = b_m$ .

Pay attention to the word “if”! Not every system of equations has a solution. For example, the system

$$\begin{cases} x_1 + x_2 + \dots + x_{100} = 0, \\ x_1 + x_2 + \dots + x_{100} = 1, \end{cases}$$



**Fig. 1.1** The intersection of two lines



of two equations in one hundred unknowns clearly cannot have any solution.

**Definition 1.5** A system possessing at least one solution is said to be *consistent*, while a system with no solutions is called *inconsistent*. If a system is consistent and has only one solution, then it is said to be *definite*, and if it has more than one solution, it is *indefinite*.

A definite system is also called *uniquely determined*, since it has precisely one solution.

Definite systems of equations are encountered frequently, for instance when from external considerations it is clear that there is only one solution. For example, suppose we wish to find the unique point lying on the lines defined by the equations  $x = y$  and  $x + y = 1$ ; see Fig. 1.1. It is clear that these lines are not parallel and therefore have exactly one point of intersection. This means that the system consisting of the equations of these two lines is definite. It is easy to find its unique solution by a simple calculation. To do so, one may substitute the condition  $y = x$  into the second equation. This yields  $2x = 1$ , that is,  $x = 1/2$ , and since  $y = x$ , we have also  $y = 1/2$ .

The reader has almost certainly encountered indefinite systems in secondary school, for example, the system

$$\begin{cases} x - 2y = 1, \\ 3x - 6y = 3. \end{cases} \quad (1.11)$$

It is obvious that the second equation is obtained by multiplying the first equation by 3. Therefore, the system is satisfied by all  $x$  and  $y$  that satisfy the first equation. From the first equation, we obtain  $2y = x - 1$ , or equivalently,  $y = (x - 1)/2$ . We can now choose an arbitrary value for  $x$  and obtain the corresponding value  $y = (x - 1)/2$ . Our system thus has infinitely many solutions and is therefore indefinite.

We have now seen examples of the following types of systems of equations:

- (a) having no solutions (inconsistent),
- (b) having a unique solution (consistent and definite),
- (c) having infinitely many solutions (for example, system (1.11)).

Let us show that these three cases are the only possibilities.

**Theorem 1.6** *If a system of linear equations is consistent and indefinite, then it has infinitely many solutions.*

*Proof* By the hypothesis of the theorem, we have a system of linear equations that is consistent and that contains more than one solution. This means that it has at least two distinct solutions:  $\mathbf{c}$  and  $\mathbf{d}$ . We shall now construct an infinite number of solutions.

To do so, we consider, for an arbitrary number  $p$ , the row  $\mathbf{r} = p\mathbf{c} + (1 - p)\mathbf{d}$ . We shall show first of all that the row  $\mathbf{r}$  is also a solution. We suppose our system to be written in the form (1.10). Then we must show that  $F_i(\mathbf{r}) = b_i$  for all  $i = 1, \dots, m$ . Using properties (1.8) and (1.9), we obtain

$$F_i(\mathbf{r}) = F_i(p\mathbf{c} + (1 - p)\mathbf{d}) = pF_i(\mathbf{c}) + (1 - p)F_i(\mathbf{d}) = pb_i + (1 - p)b_i = b_i,$$

since  $\mathbf{c}$  and  $\mathbf{d}$  are solutions of the system of equations (1.10), that is,  $F_i(\mathbf{c}) = F_i(\mathbf{d}) = b_i$  for all  $i = 1, \dots, m$ .

It remains to verify that for different numbers  $p$  we obtain different solutions. Then we will have shown that we have infinitely many of them. Let us suppose that two different numbers  $p$  and  $p'$  yield the same solution  $p\mathbf{c} + (1 - p)\mathbf{d} = p'\mathbf{c} + (1 - p')\mathbf{d}$ . We observe that we can operate on rows just as on numbers in that we can move terms from one side of the equation to the other and remove a common factor from the terms inside parentheses. This is justified because we defined operations on rows in terms of operations on the numbers that constitute them. As a result, we obtain the relation  $(p - p')\mathbf{c} = (p - p')\mathbf{d}$ . Since by assumption,  $p \neq p'$ , we can cancel the factor  $p - p'$ . On doing so, we obtain  $\mathbf{c} = \mathbf{d}$ , but by hypothesis,  $\mathbf{c}$  and  $\mathbf{d}$  were distinct solutions. From this contradiction, we conclude that every choice of  $p$  yields a distinct solution.  $\square$

## 1.2 Gaussian Elimination

Our goal now is to demonstrate a method of determining to which of the three types mentioned in the previous section a given system of linear equations belongs, that is, whether it is consistent, and if so, whether it is definite. If it is consistent and definite, then we would like to find its unique solution, and if it is consistent and indefinite, then we want to write down its solutions in some useful form. There exists a simple method that is effective in each concrete situation. It is called *Gaussian elimination*, or Gauss's method, and we now present it. We are going to be dealing here with proof by induction. That is, beginning with the simplest case, with  $m = 1$  equations, we then move on to the case  $m = 2$ , and so on, so that in considering the general case of a system of  $m$  linear equations, we shall assume that we have proved the result for systems with fewer than  $m$  equations.

The method of Gaussian elimination is based on the idea of replacing the given system of linear equations with another system having the same solutions. Let us

consider along with system (1.10) another system of linear equations in the same number of unknowns:

$$\begin{cases} G_1(\mathbf{x}) = f_1, \\ \dots \\ G_l(\mathbf{x}) = f_l, \end{cases} \quad (1.12)$$

where  $G_1(\mathbf{x}), \dots, G_l(\mathbf{x})$  are some other linear functions in  $n$  unknowns. The system (1.12) is said to be *equivalent* to system (1.10) if both systems have exactly the same solutions, that is, any solution of system (1.10) is also a solution of system (1.12), and vice versa.

The idea behind Gaussian elimination is to use certain *elementary row operations* on the system that replace a system with an equivalent but simpler system for which the answers to the questions about solutions posed above are obvious.

**Definition 1.7** An *elementary row operation of type I* on system (1.3) or (1.10) consists in the transposition of two rows. So that there will be no uncertainty about what we mean, let us be precise: under this row operation, all the equations of the system other than the  $i$ th and the  $k$ th are left unchanged, while the  $i$ th and  $k$ th exchange places.

Thus the number of elementary row operations of type I is equal to the number of pairs  $i, k, i \neq k$ , that is, the number of combinations of  $m$  things taken 2 at a time.

**Definition 1.8** An *elementary row operation of type II* consists in the replacement of the given system by another in which all equations except the  $i$ th remain as before, and to the  $i$ th equation is added  $c$  times the  $k$ th equation. As a result, the  $i$ th equation in system (1.3) takes the form

$$(a_{i1} + ca_{k1})x_1 + (a_{i2} + ca_{k2})x_2 + \dots + (a_{in} + ca_{kn})x_n = b_i + cb_k. \quad (1.13)$$

An elementary row operation of type II depends on the choice of the indices  $i$  and  $k$  and the number  $c$ , and so there are infinitely many row operations of this type.

**Theorem 1.9** *Application of an elementary row operation of type I or II results in a system that is equivalent to the original one.*

*Proof* The assertion is completely obvious in the case of an elementary row operation of type I: whatever solutions a system may have cannot depend on the numbering of its equations (that is, on the ordering of the system (1.3) or (1.10)). We could even not number the equations at all, but write each of them, for example, on a separate piece of paper.

In the case of an elementary row operation of type II, the assertion is also fairly obvious. Any solution  $\mathbf{c} = (c_1, \dots, c_n)$  of the first system after the substitution satisfies all the equations obtained under this elementary row operation except possibly

the  $i$ th, simply because they are identical to the equations of the original system. It remains to settle the question for the  $i$ th equation. Since  $\mathbf{c}$  was a solution of the original system, we have the following equalities:

$$\begin{cases} a_{i1}c_1 + a_{i2}c_2 + \cdots + a_{in}c_n = b_i, \\ a_{k1}c_1 + a_{k2}c_2 + \cdots + a_{kn}c_n = b_k. \end{cases}$$

After adding  $c$  times the second of these equations to the first, we obtain equality (1.13) for  $x_1 = c_1, \dots, x_n = c_n$ . This means that  $\mathbf{c}$  satisfies the  $i$ th equation of the new system; that is,  $\mathbf{c}$  is a solution.

It remains to prove the reverse assertion, that any solution of the system obtained by a row operation of type II is a solution of the original system. To this end, we observe that adding  $-c$  times the  $k$ th equation to equation (1.13) yields the  $i$ th equation of the original system. That is, the original system is obtained from the new system by an elementary row operation of type II using the factor  $-c$ . Thus, the previous line of argument shows that any solution of the new system obtained by an elementary row operation of type II is also a solution of the original system.  $\square$

Let us now consider Gauss's elimination method. As our first operation, let us perform on system (1.3) an elementary row operation of type I by transposing the first equation and any other in which  $x_1$  appears with a coefficient different from 0. If the first equation possesses this property, then no such transposition is necessary. Now, it can happen that  $x_1$  appears in all the equations with coefficient 0 (that is,  $x_1$  does not appear at all in the equations). In that case, we can change the numbering of the unknowns and designate by  $x_1$  some unknown that appears in some equation with nonzero coefficient. After this completely elementary transformation, we will have obtained that  $a_{11} \neq 0$ . For completeness, we should examine the extreme case in which *all* unknowns appear in *all* equations with zero coefficients. But in that case, the situation is trivial: all the equations take the form  $0 = b_i$ . If all the  $b_i$  are 0, then we have the identities  $0 = 0$ , which are satisfied for all values assigned to  $x_i$ , that is, the system is consistent and indeterminate. But if a single  $b_i$  is not equal to zero, then that  $i$ th equation is not satisfied for any values of the unknowns, and the system is inconsistent.

Now let us perform a sequence of elementary row operations of type II, adding to the second, third, and so on up to the  $m$ th equation the first equation multiplied respectively by some numbers  $c_2, c_3, \dots, c_m$  in order to make the coefficient of  $x_1$  in each of these equations equal to zero. It is clear that to do this, we must set  $c_2 = -a_{21}a_{11}^{-1}$ ,  $c_3 = -a_{31}a_{11}^{-1}$ ,  $\dots$ ,  $c_m = -a_{m1}a_{11}^{-1}$ , which is possible because we have ensured by hypothesis that  $a_{11} \neq 0$ . As a result, the unknown  $x_1$  appears in none of the equations except the first. We have thereby obtained a system that can be written in the following form:

[illegible]

Since system (1.14) was obtained from the original system (1.3) by elementary row operations, it follows from Theorem 1.3 that the two systems are equivalent, that is, the solution of an arbitrary system (1.3) has been reduced to the solution of the simpler system (1.14). That is precisely the idea behind the method of Gaussian elimination. It in fact reduces the problem to the solution of a system of  $m - 1$  equations:

[illegible]

Now if system (1.15) is inconsistent, then clearly, the larger system (1.14) is also inconsistent. If system (1.15) is consistent and we know the solution, then we can obtain all solutions of system (1.14). Namely, if  $x_2 = c_2, \dots, x_n = c_n$  is any solution of system (1.15), then we have only to substitute these values into the first equation of the system (1.14). As a result, the first equation of system (1.14) takes the form

$$a_{11}x_1 + a_{12}c_2 + \cdots + a_{1n}c_n = b_1, \quad (1.16)$$

and we have one linear equation for the remaining unknown  $x_1$ , which can be solved by the well-known formula

$$x_1 = a_{11}^{-1}(b_1 - a_{12}c_2 - \cdots - a_{1n}c_n),$$

which can be accomplished because  $a_{11} \neq 0$ . This reasoning is applicable in particular to the case  $m = 1$  (if we compare Gauss's method with the method of proof by induction, then this gives us the base case of the induction).

Thus the method of Gaussian elimination reduces the study of an arbitrary system of  $m$  equations in  $n$  unknowns to that of a system of  $m - 1$  equations in  $n - 1$  unknowns. We shall illustrate this after proving several general theorems about such systems.

**Theorem 1.10** *If the number of unknowns in a system of equations is greater than the number of equations, then the system is either inconsistent or indefinite.*

In other words, by Theorem 1.6, we know that the number of solutions of an arbitrary system of linear equations is 0, 1, or infinity. If the number of unknowns in a system is greater than the number of equations, then Theorem 1.8 asserts that the only possible number of solutions is 0 or infinity.

*Proof of Theorem 1.10* We shall prove the theorem by induction on the number  $m$  of equations in the system. Let us begin by considering the case  $m = 1$ , in which case we have a single equation:

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b_1. \quad (1.17)$$

We have  $n > 1$  by hypothesis, and if even one  $a_i$  is nonzero, then we can number the unknowns in such a way that  $a_1 \neq 0$ . We then have the case of equation (1.16). We saw that in this case, the system was consistent and indefinite.

But there remains one case to consider, that in which  $a_i = 0$  for all  $i = 1, \dots, n$ . If in this case  $b_1 \neq 0$ , then clearly we have an inconsistent “system” (consisting of a single inconsistent equation). If, however,  $b_1 = 0$ , then a solution consists of an arbitrary sequence of numbers  $x_1 = c_1, x_2 = c_2, \dots, x_n = c_n$ , that is, the “system” (consisting of the equation  $0 = 0$ ) is indefinite.

Now let us consider the case of  $m > 1$  equations. We employ the method of Gaussian elimination. That is, after writing down our system in the form (1.3), we transform it into the equivalent system (1.14). The number of unknowns in the system (1.15) is  $n - 1$ , and therefore larger than the number of equations  $m - 1$ , since by the hypothesis of the theorem,  $n > m$ . This means that the hypothesis of the theorem is satisfied for system (1.15), and by induction, we may conclude that the theorem is valid for this system. If system (1.15) is inconsistent, then all the more so is the larger system (1.14). If it is indefinite, that is, has more than one solution, then in the initial system there will be more than one solution; that is, system (1.3) will be indefinite.  $\square$

Let us now focus attention on an important special case of Theorem 1.10. A system of linear equations is said to be *homogeneous* if all the constant terms are equal to zero, that is, in (1.3), we have  $b_1 = \cdots = b_m = 0$ . A homogeneous system is always consistent: it has the obvious solution  $x_1 = \cdots = x_n = 0$ . Such a solution is called a *null* solution. We obtain the following corollary to Theorem 1.10.

**Corollary 1.11** *If in a homogeneous system, the number of unknowns is greater than the number of equations, then the system has a solution that is different from the null solution.*

If we denote (as we have been doing) the number of unknowns by  $n$  and the number of equations by  $m$ , then we have considered the case  $n > m$ . Theorem 1.10 asserts that for  $n > m$ , a system of linear equations cannot have a unique solution. Now we shall move on to consider the case  $n = m$ . We have the following rather surprising result.

**Theorem 1.12** *If in a system of linear equations, the number of unknowns is equal to the number of equations, then the property of having a unique solution depends only on the values of the coefficients and not on the values of the constant terms.*

*Proof* The result is easily obtained by Gaussian elimination. Let the system be written in the form (1.3), with  $n = m$ . Let us deal separately with the case that all the co-

efficients  $a_{ik}$  are zero (in all equations), in which case the system cannot be uniquely determined regardless of the constants  $b_i$ . Indeed, if even a single  $b_i$  is not equal to zero, then the  $i$ th equation gives an inconsistent equation; and if all the  $b_i$  are zero, then every choice of values for the  $x_i$  gives a solution. That is, the system is indefinite.

Let us prove Theorem 1.12 by induction on the number of equations ( $m = n$ ). We have already considered the case in which all the coefficients  $a_{ik}$  are equal to zero. We may therefore assume that among the coefficients  $a_{ik}$ , some are nonzero and the system can be written in the equivalent form (1.14). But the solutions to (1.14) are completely determined by system (1.15). In system (1.15), again the number of equations is equal to the number of unknowns (both equal to  $m - 1$ ). Therefore, reasoning by induction, we may assume that the theorem has been proved for this system. However, we have seen that consistency or definiteness of system (1.14) was the same as that for system (1.15). In conclusion, it remains to observe that the coefficients  $a'_{ik}$  of system (1.15) are obtained from the coefficients of system (1.3) by the formulas

$$a'_{2k} = a_{2k} - \frac{a_{21}}{a_{11}} a_{1k}, \quad a'_{3k} = a_{3k} - \frac{a_{31}}{a_{11}} a_{1k}, \quad \dots, \quad a'_{mk} = a_{mk} - \frac{a_{m1}}{a_{11}} a_{1k}.$$

Thus the question of a unique solution is determined by the coefficients of the original system (1.3).  $\square$

Theorem 1.12 can be reformulated as follows: if the number of equations is equal to the number of unknowns and the system has a unique solution for certain values of the constant terms  $b_i$ , then it has a unique solution for all possible values of the constant terms. In particular, as a choice of these “certain” values we may take all the constants to be zero. Then we obtain a system with the same coefficients for the unknowns as in system (1.3), but now the system is homogeneous. Such a system is called the homogeneous system *associated with* system (1.3). We see, then, that if the number of equations is equal to the number of unknowns, then the system has a unique solution if and only if its associated system has a unique solution. Since a homogeneous system always has the null solution, its having a unique solution is equivalent to the absence of nonnull solutions, and we obtain the following result.

**Corollary 1.13** *If in a system of linear equations, the number of equations is equal to the number of unknowns, then it has a unique solution if and only if its associated homogeneous system has no solutions other than the null solution.*

This result is unexpected, since from the *absence* of a solution different from the null solution, it derives the *existence* and uniqueness of the solution to a different system (with different constant terms). In functional analysis, this result is called the *Fredholm alternative*.<sup>1</sup>

---

<sup>1</sup>More precisely, the Fredholm alternative comprises several assertions, one of which is analogous to the one established above.

In order to focus on the theory behind the Gaussian method, we emphasized its “inductive” character: it reduces the study of a system of linear equations to an analogous system, but with fewer equations and unknowns. It is understood that in concrete examples, we must repeat the process, using this latter system and continuing until the process stops (that is, until it can no longer be applied). Now let us make clear for ourselves the form that the resulting system will take.

When we transform system (1.3) into the equivalent system (1.14), it can happen that not all the unknowns  $x_2, \dots, x_n$  enter into the corresponding system (1.15), that is, some of the unknowns may have zero coefficients in all the equations. Moreover, it was not easy to surmise this from the original system (1.3). Let us denote by  $k$  the first index of the unknown that appears with coefficients different from zero in at least one equation of system (1.15). It is clear that  $k > 1$ . We can now apply the same operations to this system. As a result, we obtain the following equivalent system:

$$\left\{ \begin{array}{l} a_{11}x_1 + \dots + a_{1n}x_n = b_1, \\ \quad a'_{2k}x_k + \dots + a'_{2n}x_n = b'_2, \\ \quad \quad a''_{3l}x_l + \dots + a''_{3n}x_n = b''_3, \\ \quad \quad \quad \dots \\ \quad \quad \quad \dots \\ \quad \quad \quad a''_{ml}x_l + \dots + a''_{mn}x_n = b''_m. \end{array} \right.$$

Here we have already chosen  $l > k$  such that in the system obtained by removing the first two equations, the unknown  $x_l$  appears with a coefficient different from zero in at least one equation. In this case we will have  $a_{11} \neq 0$ ,  $a'_{2k} \neq 0$ ,  $a'_{3l} \neq 0$ , and  $l > k > 1$ .

We shall repeat this process as long as possible. When shall we be forced to stop? We stop after having applied the elementary operations up to the point (let us say the  $r$ th equation in which  $x_s$  is the first unknown with nonzero coefficient) at which we have reduced to zero all the coefficients of all subsequent unknowns in all the remaining equations, that is, from the  $(s + 1)$ st to the  $n$ th. The system then has the following form:

$$\left\{ \begin{array}{l} \bar{a}_{11}x_1 + \dots + \bar{a}_{1n}x_n = \bar{b}_1, \\ \quad \bar{a}_{2k}x_k + \dots + \bar{a}_{2n}x_n = \bar{b}_2, \\ \quad \quad \bar{a}_{3l}x_l + \dots + \bar{a}_{3n}x_n = \bar{b}_3, \\ \quad \quad \quad \dots \\ \quad \quad \quad \dots \\ \quad \quad \quad \bar{a}_{rs}x_s + \dots + \bar{a}_{rn}x_n = \bar{b}_r, \\ \quad \quad \quad \quad \quad 0 = \bar{b}_{r+1}, \\ \quad \quad \quad \quad \quad \dots \\ \quad \quad \quad \quad \quad 0 = \bar{b}_m. \end{array} \right. \quad (1.18)$$

Here  $1 < k < l < \dots < s$ .



It can happen that  $r = m$ , and therefore, there will be no equations of the form  $0 = \bar{b}_i$  in system (1.18). But if  $r < m$ , then it can happen that  $\bar{b}_{r+1} = 0, \dots, \bar{b}_m = 0$ , and it can finally be the case that one of the numbers  $\bar{b}_{r+1}, \dots, \bar{b}_m$  is different from zero.

**Definition 1.14** System (1.18) is said to be in (row) *echelon* form. The same terminology is applied to the matrix of such a system.

**Theorem 1.15** Every system of linear equations is equivalent to a system in echelon form (1.18).

*Proof* Since we transformed the initial system into the form (1.18) using a sequence of elementary row operations, it follows from Theorem 1.9 that system (1.18) is equivalent to the initial system.  $\square$

Since any system of the form (1.3) is equivalent to system (1.18) in echelon form, questions about consistency and definiteness of systems can be answered by studying systems in echelon form.

Let us begin with the question of consistency. It is clear that if system (1.18) contains equations  $0 = \bar{b}_k$  with  $\bar{b}_k \neq 0$ , then such a system is inconsistent, since the equality  $0 = \bar{b}_k$  cannot be satisfied by any values of the unknowns. Let us show that if there are no such equations in system (1.18), then the system is consistent. Thus we now assume that in system (1.18), the last  $m - r$  equations have been converted into the identities  $0 \equiv 0$ .

Let us call the unknowns  $x_1, x_k, x_l, \dots, x_s$  that begin the first, second, third,  $\dots$ ,  $r$ th equations of system (1.18) *principal*, and the rest of the unknowns (if there are any) we shall call *free*. Since every equation in system (1.3) begins with its own principal unknown, the number of principal unknowns is equal to  $r$ . We recall that we have assumed  $\bar{b}_{r+1} = \dots = \bar{b}_m = 0$ .

Let us assign arbitrary values to the free unknowns and substitute them in the equations of system (1.18). Since the  $r$ th equation contains only one principal unknown  $x_s$ , and that with the coefficient  $\bar{a}_{rs}$ , which is different from zero, we obtain for  $x_s$  one equation in one unknown, which has a unique solution. Substituting this solution for  $x_s$  into the equation above it, we obtain for that equation's principal unknown again one equation in one unknown, which also has a unique solution. Continuing in this way, moving from bottom to top in system (1.18), we see that the values of the principal unknowns are determined uniquely for an arbitrary assignment of the free unknowns. We have thus proved the following theorem.

**Theorem 1.16** For a system of linear equations to be consistent, it is necessary and sufficient, after it has been brought into echelon form, that there be no equations of the form  $0 = \bar{b}_k$  with  $\bar{b}_k \neq 0$ . If this condition is satisfied, then it is possible to assign arbitrary values to the free unknowns, while the values of the principal unknowns—for each given set of values for the free unknowns—are determined uniquely from the system.

Let us now explain when a system will be definite on the assumption that the condition of consistency that we have been investigating is satisfied. This question is easily answered on the basis of Theorem 1.16. Indeed, if there are free unknowns in system (1.18), then the system is certainly not definite, since we may give an arbitrary assignment to each of the free unknowns, and by Theorem 1.16, the assignment of principal unknowns is then determined by the system. On the other hand, if there are no free unknowns, then all the unknowns are principal. By Theorem 1.16, they are uniquely determined by the system, which means that the system is definite. Consequently, a necessary and sufficient condition for definiteness is that there be no free unknowns in system (1.18). This, in turn, is equivalent to all unknowns in the system being principal. But that, clearly, is equivalent to the equality  $r = n$ , since  $r$  is the number of principal unknowns and  $n$  is the total number of unknowns. Thus we have proved the following assertion.

**Theorem 1.17** *For a consistent system (1.3) to be definite, it is necessary and sufficient that for system (1.18), after it has been brought into echelon form, we have the equality  $r = n$ .*

**Remark 1.18** Any system of  $n$  equations in  $n$  unknowns (that is, with  $m = n$ ) brought into echelon form can be written in the form

$$\left\{ \begin{array}{l} \bar{a}_{11}x_1 + \bar{a}_{12}x_2 + \cdots + \bar{a}_{1n}x_n = \bar{b}_1, \\ \quad \bar{a}_{22}x_2 + \cdots + \bar{a}_{2n}x_n = \bar{b}_2, \\ \quad \quad \quad \cdots \quad \quad \quad \cdots \\ \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \cdots \\ \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \bar{a}_{nn}x_n = \bar{b}_n \end{array} \right. \quad (1.19)$$

(however, not every system of the form (1.19) is in echelon form, since some of the  $\bar{a}_{ii}$  can be zero). Indeed, the form (1.19) indicates that in the system, the  $k$ th equation does not depend on the unknowns  $x_i$  for  $i < k$ , and this condition is automatically satisfied for a system in echelon form.

A system in the form (1.19) is said to be in *upper triangular form*. The same terminology is applied to the matrix of system (1.19).

From this observation, we can state Theorem 1.15 in a different form for the case  $m = n$ . The condition  $r = n$  means that all the unknowns  $x_1, x_2, \dots, x_n$  are principal, and that means that in system (1.19), the coefficients satisfy  $\bar{a}_{11} \neq 0, \dots, \bar{a}_{nn} \neq 0$ . This proves the following corollary.

**Corollary 1.19** *System (1.3) in the case  $m = n$  is consistent and determinate if and only if after being brought into echelon form, we obtain the upper triangular system (1.19) with coefficients  $\bar{a}_{11} \neq 0, \bar{a}_{22} \neq 0, \dots, \bar{a}_{nn} \neq 0$ .*

We see that this condition is independent of the constant terms, and we thereby obtain another proof of Theorem 1.12 (though it is based on the same idea of the method of Gaussian elimination).



is  $n + 1$  (the numeration begins here not with the usual  $a_1$ , but with  $a_0$ ). The numbers 1 and  $c_i^k$  are the coefficients of the unknowns, and  $k_1, \dots, k_r$  are the constant terms.

If  $r = n + 1$ , then we are in the situation of Theorem 1.12 and its corollary. Therefore, for  $r = n + 1$ , the interpolation problem has a solution, and a unique one, if and only if the associated system (1.20) has only the null solution. This associated system can be written in the form

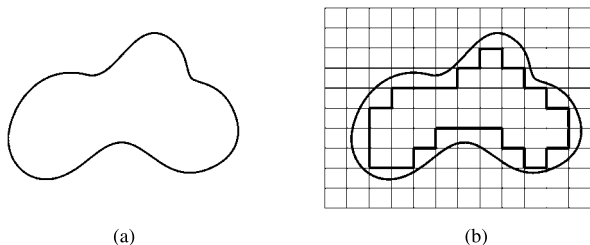
$$\begin{cases} f(c_1) = 0, \\ f(c_2) = 0, \\ \dots \\ f(c_r) = 0. \end{cases} \quad (1.21)$$

A number  $c$  for which  $f(c) = 0$  is called a *root* of the polynomial  $f$ . A simple theorem of algebra (a corollary of what is known as Bézout's theorem) states that a polynomial cannot have more distinct roots than its degree (except in the case that all the  $a_i$  are equal to zero, in which case the degree is undefined). This means (if the numbers  $c_i$  are distinct, which is a natural assumption) that for  $r = n + 1$ , equations (1.21) can be satisfied only if all the  $a_i$  are zero. We obtain that under these conditions, system (1.20) (that is, the interpolation problem) has a solution, and the solution is unique. We note that it is not particularly difficult to obtain an explicit formula for the coefficients of the polynomial  $f$ . This will be done in Sects. 2.4 and 2.5.

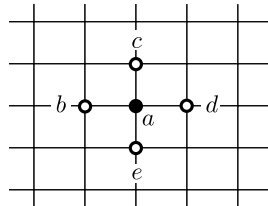
The following example is somewhat more difficult.

*Example 1.21* Many questions in physics (such as the distribution of heat in a solid body if a known temperature is maintained on its surface, or the distribution of electric charge on a body if a known charge distribution is maintained on its surface, and so on) lead to a single differential equation, called the Laplace equation. It is a partial differential equation, which we do not need to describe here. It suffices to mention one consequence, called the *mean value property*, according to which the value of the unknown quantity (satisfying the Laplace equation) is equal at every point to the arithmetic mean of its values at “nearby” points. We need not make precise here just what we mean by “nearby points” (suffice it to say that there are infinitely many of them, and this property is defined in terms of the integral). We will, however, present a method for an approximate solution of the Laplace equation. Solely for the purpose of simplifying the presentation, we shall consider the two-dimensional case instead of the three-dimensional situation described above. That is, instead of a three-dimensional body and its surface, we shall examine a two-dimensional figure and its boundary; see Fig. 1.3(a). To construct an approximate solution in the plane, we form a lattice of identical small squares (the smaller the squares, the better the approximation), and the contour of the figure will be replaced by the closest approximation to it consisting of sides of the small squares; see Fig. 1.3(b).

**Fig. 1.3** Constructing an approximate solution to the Laplace equation



**Fig. 1.4** The “nearby vertices” to  $a$  are the points  $b, c, d, e$



We examine the values of the unknown quantity (temperature, charge, etc.) only at the vertices of the small squares. Now the concept of “nearby points” acquires an unambiguous meaning: each vertex of a square of the lattice has exactly four nearby points, namely the “nearby” vertices. For example, in Fig. 1.4, the point  $a$  has nearby vertices  $b, c, d, e$ .

We consider as given some quantities  $x_a$  for all the vertices  $a$  of the squares intersecting the boundary (the thick straight lines in Fig. 1.3(b)), and we seek such values for the vertices of the squares located inside this contour. Now an approximate analogue of the mean value property for the point  $a$  of Fig. 1.4 is the relationship

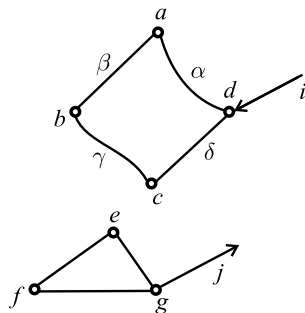
$$x_a = \frac{x_b + x_c + x_d + x_e}{4}. \quad (1.22)$$

There are thus as many unknowns as there are vertices inside the contour, and to each such vertex there corresponds an equation of type (1.22). This means that we have a system of linear equations in which the number of equations is equal to the number of unknowns. If one of the vertices  $b, c, d, e$  is located on the contour, then the corresponding quantity, one of  $x_b, x_c, x_d, x_e$ , must be assigned, and equation (1.22) in this case is inhomogeneous. An assertion from the theory of linear equations that we shall prove is that regardless of how we assign values on the boundary of the figure, the associated system of linear equations always has a unique solution.

We clearly find ourselves in the situation of Corollary 1.13, and so it suffices to verify that the homogeneous system associated with ours has only the null solution. The associated homogeneous system corresponds to the case in which all the values on the boundary of the figure are equal to zero. Let us suppose that it has a solution  $x_1, \dots, x_N$  (where  $N$  is the number of equations) that is not the null solution. If among the numbers  $x_i$  there is at least one that is positive, then let us denote by  $x_a$  the largest such number. Then equation (1.22) (in which any of  $x_b, x_c, x_d, x_e$  will



**Fig. 1.7** Decomposable network



rent  $i$  enters, while at another point, current  $j$  exits. A uniform current flows due to the homogeneity of each conductor.

We shall designate the conductors by the Greek letters  $\alpha, \beta, \gamma, \dots$ , and the strength of the current in conductor  $\alpha$  by  $i_\alpha$ . Knowing the current  $i$ , we would like to find the currents  $i_\alpha, i_\beta, i_\gamma, \dots$  for all the conductors in the network  $\alpha, \beta, \gamma, \dots$ , and the current  $j$ . We shall denote the nodes of the network by  $a, b, c, \dots$ .

We need to make one additional refinement here. Since the current in a conductor flows in a particular direction, it makes sense to indicate the direction with a sign. This choice is arbitrary for each conductor, and we designate the direction by an arrow. The nodes joined by a conductor are called its *beginning* and *end*, and the arrow points from the beginning of the conductor to the end. The beginning of the conductor  $\alpha$  will be denoted by  $\alpha'$ , and the end will be denoted by  $\alpha''$ . The current  $i_\alpha$  will be considered positive if it flows in the direction of the arrow, and will be considered negative otherwise. We shall say that the current  $i_\alpha$  flows out of node  $a$  (flows into node  $a$ ) if there is a conductor  $\alpha$  with beginning (end) node  $a$ . For example, in Fig. 1.6, the current  $i_\alpha$  flows out of  $a$  and flows into  $b$ ; thus according to our notation,  $\alpha' = a$  and  $\alpha'' = b$ .

We shall assume further that the network in question satisfies the following natural condition: Two arbitrary nodes  $a$  and  $b$  can be connected by some set of nodes  $c_1, \dots, c_n$  in such a way that each of the pairs  $a, c_1; c_1, c_2; \dots; c_{n-1}, c_n; c_n, b$  are connected by a conductor. We shall call this property of the network *connectedness*. A network not satisfying this condition can be decomposed into a number of subnetworks each of whose nodes are not connected to any nodes of any other subnetwork (Fig. 1.7). We may then consider each subnetwork individually.

A collection of nodes  $a_1, \dots, a_n$  connecting conductors  $\alpha_1, \dots, \alpha_n$  such that conductor  $\alpha_1$  connects nodes  $a_1$  and  $a_2$ , conductor  $\alpha_2$  connects nodes  $a_2$  and  $a_3$ , ..., conductor  $\alpha_{n-1}$  connects nodes  $a_{n-1}$  and  $a_n$ , and conductor  $\alpha_n$  connects nodes  $a_n$  and  $a_1$  is called a closed circuit. For example, in Fig. 1.6, it is possible to select as a closed circuit nodes  $a, b, c, d, h$  and conductors  $\alpha, \beta, \gamma, \xi, \eta$ , or else, for example, nodes  $e, g, h, d$  and conductors  $\mu, \vartheta, \xi, \delta$ . The distribution of current in the closed circuit is determined by two well-known laws of physics: Kirchhoff's laws.

*Kirchhoff's first law* applies to each node of a network and asserts that the sum of the currents flowing into a node is equal to the sum of the currents flowing out it. More precisely, the sum of the currents in the conductors that have node  $a$  at their

end is equal to the sum of the currents in the conductors for which node  $a$  is the beginning. This can be expressed by the following formula:

$$\sum_{\alpha'=a} i_{\alpha} - \sum_{\beta''=a} i_{\beta} = 0 \quad (1.23)$$

for every node  $a$ . For example, in Fig. 1.6, for the node  $e$  we obtain the equation

$$i_{\varepsilon} - i_{\delta} - i_{\lambda} - i_{\mu} = 0.$$

*Kirchhoff's second law* applies to an arbitrary closed circuit consisting of conductors in a network. Namely, if the conductors  $\alpha_l$  form a circuit  $C$ , then with a direction of such a circuit having been assigned, the law is expressed by the equation

$$\sum_{\alpha_l \in C} \pm p_{\alpha_l} i_{\alpha_l} = 0, \quad (1.24)$$

where  $p_{\alpha_l}$  is the *resistance* of the conductor  $\alpha_l$  (which is always a positive number, since the conductors are homogeneous), and where the plus sign is taken if the selected direction of the conductor (indicated by an arrow) coincides with the direction of the current in the circuit, and the minus sign is taken if it is opposite to the direction of the current. For example, for the closed circuit  $C$  with nodes  $e, g, h, d$  as shown in Fig. 1.6 and with the indicated direction of the circuit, Kirchhoff's law gives the equation

$$-p_{\mu} i_{\mu} + p_{\vartheta} i_{\vartheta} - p_{\xi} i_{\xi} + p_{\delta} i_{\delta} = 0. \quad (1.25)$$

We thereby obtain a system of linear equations in which the unknowns are  $i_{\alpha}, i_{\beta}, i_{\gamma}, \dots$  and  $j$ . Such a system of equations is encountered in a number of problems, such as the allocation of loads in a transport network and the distribution of water in a system of conduits.

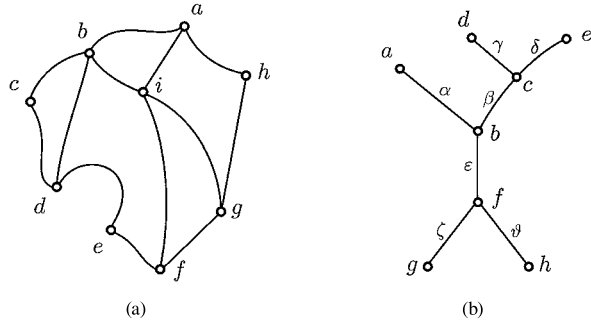
Our goal is now to show that the system of equations thus obtained (for the given network and currents  $i$ ) has a unique solution.

First, we observe that the outflowing current  $j$  is equal to  $i$ . This is obvious from physical considerations, but we must derive it from the equations of Kirchhoff's law. To this end, let us collect all equations (1.23) for Kirchhoff's first law for all nodes  $a$  of our network. How often do we encounter conductor  $\alpha$  in the obtained equation? We encounter it once when we examine the equation corresponding to the node  $a = \alpha'$ , and another time for  $a = \alpha''$ . Furthermore, the current  $i_{\alpha}$  enters into the two equations with opposite signs, which means that they cancel. All that will remain in the resulting equation is the current  $i$  (for the point into which the current flows) and  $-j$  (for the point where the current flows out). This yields the equation  $i - j = 0$ , that is,  $i = j$ .

Now let us note that not all the equations (1.24) corresponding to Kirchhoff's second law are independent. We shall call a closed circuit  $\alpha_1, \dots, \alpha_n$  a *cell* if every pair of its nodes is connected only by a conductor from among  $\alpha_1, \dots, \alpha_n$  and by no others. Every closed circuit can be decomposed into a number of cells. For



**Fig. 1.8** Circuits for the proof of Euler's theorem



example, in Fig. 1.6, the circuit  $C$  with nodes  $e, g, h, d$  and conductors  $\mu, \vartheta, \xi, \delta$  can be decomposed into two cells: one with nodes  $e, g, h$  and conductors  $\mu, \vartheta, \lambda$ , and the other with nodes  $e, h, d$  and conductors  $\lambda, \xi, \delta$ . In this case, equation (1.24) corresponding to the circuit is the sum of the equations corresponding to the individual cells (with a proper choice of directions for the circuits). For example, equation (1.25) for the circuit  $C$  with nodes  $e, g, h, d$  is the sum of equations

$$-p_{\mu}i_{\mu} + p_{\vartheta}i_{\vartheta} + p_{\lambda}i_{\lambda} = 0, \quad -p_{\lambda}i_{\lambda} - p_{\xi}i_{\xi} + p_{\delta}i_{\delta} = 0,$$

corresponding to the cells with nodes  $e, g, h$  and  $e, h, d$ .

Thus, we can restrict our attention to equations of the cells of the network. Let us prove, then, that in the entire system of equations (1.23) and (1.24) corresponding to Kirchhoff's first and second laws, the number of equations will be equal to the number of unknowns. We shall denote by  $N_{\text{cell}}$ ,  $N_{\text{cond}}$ , and  $N_{\text{node}}$  the numbers of cells, conductors, and nodes of the network. The number of unknowns  $i_{\alpha}$  and  $j$  is equal to  $N_{\text{cond}} + 1$ . Each cell and each node contributes one equation. This means that the number of equations is equal to  $N_{\text{cell}} + N_{\text{node}}$ , and we need to prove the equality

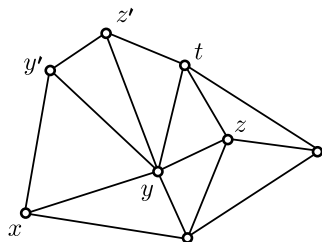
$$N_{\text{cell}} + N_{\text{node}} = N_{\text{cond}} + 1. \quad (1.26)$$

This is a familiar equality. It comes from topology and is known as *Euler's theorem*. It is very easy to prove, as we shall now demonstrate.

Let us make the important observation that our network is located in the plane: the conductors do not have to be straight line segments, but they are required to be nonintersecting curves in the plane. We shall use induction on the number of cells. Let us delete the “outer” side of one of the “external” cells (for example, side  $(b, c, d)$  in Fig. 1.8(a)). In this case, the number of cells  $N_{\text{cell}}$  is reduced by 1.

If in the “deleted” side there were  $k$  conductors, then the number  $N_{\text{cond}}$  will decrease by  $k$ , while the number  $N_{\text{node}}$  will decrease by  $k - 1$ . Altogether, the number  $N_{\text{cell}} - N_{\text{cond}} + N_{\text{node}} - 1$  does not change. In this process, the property of connectedness is not destroyed. Indeed, any two nodes of the initial network can be connected by the sequence of nodes  $c_1, \dots, c_n$ . If even part of this sequence consisted of vertices of the “deleted” sides of our cell, then we could replace them with the sequence of nodes of its “nondeleted” sides.

**Fig. 1.9** Closed circuit containing nodes  $x$  and  $t$



This process reduces the proof to the case  $N_{\text{cell}} = 0$ , that is, to a network that does not contain a closed circuit. We now must prove that for such a network,  $N_{\text{node}} - N_{\text{cond}} = 1$ . We now use induction on the number  $N_{\text{cond}}$ . Let us remove any “external” conductor at least one end of which is not the end of another conductor (for example, the conductor  $\alpha$  in Fig. 1.8(b)). Then both numbers  $N_{\text{cond}}$  and  $N_{\text{node}}$  are reduced by 1, and the number  $N_{\text{cond}} - N_{\text{node}}$  remains unchanged. We may easily convince ourselves that in this case, the property of connectedness is again preserved. As a result, we arrive at the case  $N_{\text{cond}} = 0$  but  $N_{\text{node}} > 0$ . Since the network must be connected, we have  $N_{\text{node}} = 1$ , and it is clear that we have the equality  $N_{\text{node}} - N_{\text{cond}} = 1$ .

We now note an important property of networks satisfying relationship (1.24) that emerges from Kirchhoff’s second law (for given currents  $i_\alpha$ ). With each node  $a$  one can associate a number  $r_a$  such that for an arbitrary conductor  $\alpha$  beginning at  $a$  and ending at  $b$ , the following equation is satisfied:

$$p_\alpha i_\alpha = r_a - r_b. \quad (1.27)$$

To determine these numbers  $r_\alpha$ , we shall choose some node  $x$  and assign to it the number  $r_x$  arbitrarily. Then for each node  $y$  connected to  $x$  by some conductor  $\alpha$ , we set  $r_y = r_x - p_\alpha i_\alpha$  if  $x$  is at the beginning of  $\alpha$  and  $y$  at the end, and  $r_y = r_x + p_\alpha i_\alpha$  in the opposite case. Then in exactly the same way, we determine the number  $r_z$  for each node connected by a conductor to one of the examined nodes  $x, y$ , etc. In view of the connectedness condition, we will eventually reach every node  $t$  of our network, to which we will have assigned, say, the number  $r_t$ . But it is still necessary to show that this number  $r_t$  is independent of the path by which we arrive from  $x$  to  $t$  (that is, which point we chose as  $y$ , then as  $z$ , and so on). To accomplish this, it suffices to note that a pair of distinct paths linking nodes  $x$  and  $t$  forms a closed circuit (Fig. 1.9), and the relationship that we require follows from Kirchhoff’s second law (equations (1.24)).

It is now easy to show that the system of linear equations (1.23) obtained from Kirchhoff’s first law for all nodes and from Kirchhoff’s second law (1.24) for all cells has a unique solution. To do so, it suffices, as we know, to show that the associated homogeneous system has only the null solution. This homogeneous system is obtained for  $i = j = 0$ .

Of course, “physically,” it is completely obvious that if we put no current into the network, then there will be no current in its conductors, but we must prove that this follows in particular from Kirchhoff’s laws.

To this end, consider the sum  $\sum_{\alpha} p_{\alpha} i_{\alpha}^2$ , where the sum is over all conductors of our network. Let us break the term  $p_{\alpha} i_{\alpha}^2$  into two factors:  $p_{\alpha} i_{\alpha}^2 = (p_{\alpha} i_{\alpha}) \cdot i_{\alpha}$ . We replace the first factor by  $r_a - r_b$  on the basis of relation (1.27), where  $a$  is the beginning and  $b$  the end of conductor  $\alpha$ . We obtain the sum  $\sum_{\alpha} (r_a - r_b) i_{\alpha}$ , and we collect the terms in which the first factor  $r_a$  or  $-r_b$  is associated with a particular node  $c$ . Then we can pull the number  $r_c$  outside the parentheses, and inside will remain the sum  $\sum_{\alpha'=c} i_{\alpha} - \sum_{\beta''=c} i_{\beta}$ , which is equal to zero on account of Kirchhoff's first law (1.23). We finally obtain that  $\sum_{\alpha} p_{\alpha} i_{\alpha}^2 = 0$ , and since the resistance  $p_{\alpha}$  is positive, all the currents  $i_{\alpha}$  must be equal to zero.

To conclude, we remark that networks appearing in mathematics are called *graphs*, and “conductors” become the *edges* of the graph. In the case that every edge of a graph is assigned a direction (provided with arrows, for example), the graph is then said to be *directed*. This theorem holds not for arbitrary graphs, but only for those, like the networks that we have considered in this example, that can be drawn in the plane without intersections of edges (for which we omit a precise definition). Such graphs are called *planar*.

## Chapter 2

# Matrices and Determinants

### 2.1 Determinants of Orders 2 and 3

We begin by considering a system of two equations in two unknowns:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 = b_1, \\ a_{21}x_1 + a_{22}x_2 = b_2. \end{cases}$$

In order to determine  $x_1$ , we attempt to eliminate  $x_2$  from the system. To accomplish this, it suffices to multiply the first equation by  $a_{22}$  and add to it the second equation multiplied by  $-a_{12}$ . We obtain

$$(a_{11}a_{22} - a_{21}a_{12})x_1 = b_1a_{22} - b_2a_{12}.$$

We consider the case in which  $a_{11}a_{22} - a_{21}a_{12} \neq 0$ . Then we obtain

$$x_1 = \frac{b_1a_{22} - b_2a_{12}}{a_{11}a_{22} - a_{21}a_{12}}. \quad (2.1)$$

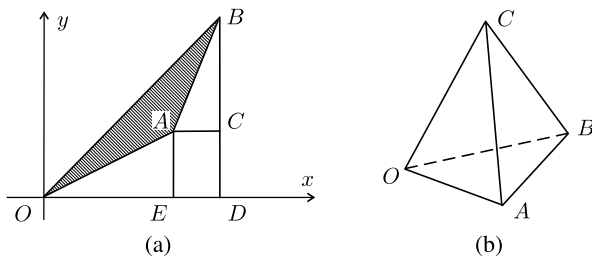
Analogously, to find the value  $x_2$ , we multiply the second equation by  $a_{11}$  and add to it the first multiplied by  $-a_{21}$ . With the same assumption ( $a_{11}a_{22} - a_{21}a_{12} \neq 0$ ), we obtain

$$x_2 = \frac{b_2a_{11} - b_1a_{21}}{a_{11}a_{22} - a_{21}a_{12}}. \quad (2.2)$$

The expression  $a_{11}a_{22} - a_{12}a_{21}$  appearing in the denominator of formulas (2.1) and (2.2) is called the *determinant* of the matrix  $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$  (it is called a determinant of order 2, or a  $2 \times 2$  determinant) and is denoted by  $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$ . Therefore, we have by definition,

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}. \quad (2.3)$$

**Fig. 2.1** Calculating (a) the area of a *triangle* and (b) the volume of a *tetrahedron*



We see that in the numerators of formulas (2.1) and (2.2) there also appears an expression of the form (2.3). Using the notation we have introduced, we can rewrite these formulas in the following form:

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}, \quad x_2 = \frac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}. \quad (2.4)$$

The expression (2.3) is useful for more than a symmetric way of writing solutions of two equations in two unknowns. It is encountered in a great number of situations, and therefore has a special name and notation. For example, consider two points  $A$  and  $B$  in the plane with respective coordinates  $(x_1, y_1)$  and  $(x_2, y_2)$ ; see Fig. 2.1(a). It is not difficult to see that the area of triangle  $OAB$  is equal to  $(x_1 y_2 - y_1 x_2)/2$ . For example, we could subtract from the area of triangle  $OBD$  the area of the rectangle  $ACDE$  and the areas of triangles  $ABC$  and  $OAE$ . We thereby obtain

$$\Delta OAB = \frac{1}{2} \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}.$$

Having in hand formulas for solutions of systems of two equations in two unknowns, we can solve some other systems. Consider, for example, the following homogeneous system of linear equations in three unknowns:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = 0, \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = 0. \end{cases} \quad (2.5)$$

We are interested in nonnull solutions of this system, that is, solutions in which at least one  $x_i$  is not equal to zero. Suppose, for example, that  $x_3 \neq 0$ . Dividing both sides by  $-x_3$  and setting  $-x_1/x_3 = y_1$ ,  $-x_2/x_3 = y_2$ , we can write system (2.5) in the form

$$\begin{cases} a_{11}y_1 + a_{12}y_2 = a_{13}, \\ a_{21}y_1 + a_{22}y_2 = a_{23}, \end{cases}$$

which is in a form we have considered. If  $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0$ , then formula (2.4) gives the expressions

$$y_1 = -\frac{x_1}{x_3} = \frac{\begin{vmatrix} a_{13} & a_{12} \\ a_{23} & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}, \quad y_2 = -\frac{x_2}{x_3} = \frac{\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}.$$

Unsurprisingly, we determined from system (2.5) not  $x_1, x_2, x_3$ , but only their mutual relationships: from such a homogeneous system, it easily follows that if  $(c_1, c_2, c_3)$  is a solution and  $p$  is an arbitrary number, then  $(pc_1, pc_2, pc_3)$  is also a solution. Therefore, we can set

$$x_1 = -\begin{vmatrix} a_{13} & a_{12} \\ a_{23} & a_{22} \end{vmatrix}, \quad x_2 = -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}, \quad x_3 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \quad (2.6)$$

and say that an arbitrary solution is obtained from this one by multiplying all the  $x_i$  by  $p$ . In order to give our solution a somewhat more symmetric form, we observe that we always have

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = -\begin{vmatrix} b & a \\ d & c \end{vmatrix}.$$

This is easily checked with the help of formula (2.3). Therefore, (2.6) can be written in the form

$$x_1 = \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}, \quad x_2 = -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}, \quad x_3 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}. \quad (2.7)$$

Formulas (2.7) give values for  $x_1, x_2, x_3$  if we cross out in turn the first, second, and third columns and then take the obtained second-order determinants with alternating signs. We recall that these formulas were obtained on the assumption that

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0.$$

It is easy to check that the assertion we have proved is valid if at least one of the three determinants appearing in (2.7) is not equal to zero. If all three determinants are zero, then, of course, formula (2.7) again gives a solution, namely the null solution, but now we can no longer assert that all solutions are obtained by multiplying by a number (indeed, this is not true).

Let us now consider the case of a system of three equations in three unknowns:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1, \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2, \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3. \end{cases}$$

We again would like to eliminate  $x_2$  and  $x_3$  from the system in order to obtain a value for  $x_1$ . To this end, we multiply the first equation by  $c_1$ , the second by  $c_2$ ,

and the third by  $c_3$  and add them. We shall therefore choose  $c_1$ ,  $c_2$ , and  $c_3$  such that in the system obtained, the terms with  $x_2$  and  $x_3$  become equal to zero. Setting the associated coefficients to zero, we obtain for  $c_1$ ,  $c_2$ , and  $c_3$  the following system of equations:

$$\begin{cases} a_{12}c_1 + a_{22}c_2 + a_{32}c_3 = 0, \\ a_{13}c_1 + a_{23}c_2 + a_{33}c_3 = 0. \end{cases}$$

This system is of the same type as (2.5). Therefore, we can use the formula (2.6) that we derived and take

$$c_1 = \begin{vmatrix} a_{22} & a_{32} \\ a_{23} & a_{33} \end{vmatrix}, \quad c_2 = -\begin{vmatrix} a_{12} & a_{32} \\ a_{13} & a_{33} \end{vmatrix}, \quad c_3 = \begin{vmatrix} a_{12} & a_{22} \\ a_{13} & a_{23} \end{vmatrix}.$$

As a result, we obtain for  $x_1$  the equation

$$\begin{aligned} & \left( a_{11} \begin{vmatrix} a_{22} & a_{32} \\ a_{23} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{32} \\ a_{13} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \right) x_1 \\ & = b_1 \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - b_2 \begin{vmatrix} a_{12} & a_{32} \\ a_{13} & a_{33} \end{vmatrix} + b_3 \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}. \end{aligned} \quad (2.8)$$

The coefficient of  $x_1$  in (2.8) is called the *determinant* of the matrix

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

and is denoted by

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

Therefore, by definition,

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}. \quad (2.9)$$

It is clear that the right-hand side of equation (2.8) is obtained from the coefficient of  $x_1$  by substituting  $a_{i1}$  for  $b_i$ ,  $i = 1, 2, 3$ . Therefore, equality (2.8) can be written in the form

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} x_1 = \begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix}.$$

We shall assume that the coefficient of  $x_1$ , that is, the determinant (2.9), is different from zero. Then we have

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}}. \quad (2.10)$$

We can easily carry out the same calculations for  $x_2$  and  $x_3$ . We obtain then the formulas

$$x_2 = \frac{\begin{vmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}}, \quad x_3 = \frac{\begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}}.$$

Just as second-order determinants express area, third-order determinants enter into a number of formulas for volume. For example, the volume of a tetrahedron with vertices at the points  $O$  (the coordinate origin) and  $A, B, C$  with coordinates  $(x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3)$  (see Fig. 2.1(b)), is equal to

$$\frac{1}{6} \begin{vmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{vmatrix}.$$

This shows that the notion of determinant that we have introduced is encountered in a number of branches of mathematics. We now return to the problem of solving systems of  $n$  linear equations in  $n$  unknowns.

It is clear that we can apply the same line of reasoning to a system consisting of four equations in four unknowns. To do so, we need to derive formulas analogous to (2.7) for the solution of a homogeneous system of three equations in four unknowns based on formula (2.9). Then to eliminate  $x_2, x_3, x_4$  in a system of four equations in four unknowns, we multiply the equations by the coefficients  $c_1, c_2, c_3, c_4$  and add. The coefficients  $c_1, c_2, c_3, c_4$  will satisfy a homogeneous system of three equations, which we are able to solve. This will give us uniquely solvable linear equations in the unknowns  $x_1, \dots, x_4$  (as in the previous cases with two and three variables, the idea is the same for any number of unknowns). We call the coefficient of the unknowns a fourth-order *determinant*. Solving the linear equations thus obtained, we arrive at formulas expressing the values of the unknowns  $x_1, \dots, x_4$ , analogous to formula (2.10). Thus it is possible to obtain solutions to systems with an arbitrarily large number of equations and with the same number of unknowns.

To derive a formula for the solution of  $n$  equations in  $n$  unknowns, we have to introduce the notion of the determinant of the  $n \times n$  square matrix

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}, \quad (2.11)$$



that is, a determinant of order  $n$ .

Our previous analysis suggests that we define the  $n \times n$  determinant by induction: For  $n = 1$ , we consider the determinant of the matrix  $(a_{11})$  to be equal to the number  $a_{11}$ , and assuming that the determinant of order  $n - 1$  has been defined, we proceed to define the determinant of order  $n$ .

Formulas (2.3) and (2.9) suggest how this should be done. In both formulas, the determinant of order  $n$  (that is, two or three) was expressed in the form of an algebraic sum of elements of the first column of matrix (2.11) (that is, of elements  $a_{11}, a_{21}, \dots, a_{n1}$ ) multiplied by determinants of order  $n - 1$ . The determinant of order  $n - 1$  by which a given element of the first column was multiplied was obtained by deleting from the original matrix the first column and the row in which the given element was located. Then the  $n$  products were added with alternating signs.

We shall give a general definition of an  $n \times n$  determinant in the following section. The sole purpose of the discussion above was to make such a definition intelligible. The formulas introduced in this section will not be used again in this book. Indeed, they will be corollaries of formulas that we shall derive for determinants of arbitrary order.

## 2.2 Determinants of Arbitrary Order

A determinant of the square  $n \times n$  matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

is a number associated with the given matrix. It is defined inductively on the number  $n$ . For  $n = 1$ , the determinant of the matrix  $(a_{11})$  is simply the number  $a_{11}$ . Suppose that we know how to compute the determinant of an arbitrary matrix of order  $(n - 1)$ . We then define the *determinant* of a square matrix  $A$  as the product

$$|A| = a_{11}D_1 - a_{21}D_2 + a_{31}D_3 - a_{41}D_4 + \cdots + (-1)^{n+1}a_{n1}D_n, \quad (2.12)$$

where  $D_k$  is the determinant of order  $(n - 1)$  obtained from the matrix  $A$  by deleting the first column and the  $k$ th row. (The reader should verify that for  $n = 2$  and  $n = 3$  we obtain the same formulas for determinants of order 2 and 3 presented in the previous section.)

Let us now introduce some useful notation and terminology. The determinant of the matrix  $A$  is denoted by

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix},$$

or simply by  $|A|$ , for short. If we delete the  $i$ th row and the  $j$ th column of the matrix  $A$  and preserve the ordering of the remaining elements, then we end up with a matrix of order  $(n - 1)$ . Its determinant is denoted by  $M_{ij}$  and is called a *minor* of the matrix  $A$ , or more precisely, the minor associated with the element  $a_{ij}$ . With this notation, (2.12) can be written in the form

$$|A| = a_{11}M_{11} - a_{21}M_{21} + a_{31}M_{31} - \cdots + (-1)^{n+1}a_{n1}M_{n1}. \quad (2.13)$$

This formula can be expressed in words thus: *The determinant of an  $n \times n$  matrix is equal to the sum of the elements of the first column each multiplied by its associated minor, where the sum is taken with alternating signs, beginning with plus.*

*Example 2.1* Suppose a particular square matrix  $A$  of order  $n$  has the property that all of its elements in the first column are equal to zero except for the element in the first row. That is,

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

Then in (2.13), all the terms except the first are equal to zero. Then formula (2.13) gives the equality

$$|A| = a_{11}|A'|, \quad (2.14)$$

where the matrix

$$A' = \begin{pmatrix} a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

is of order  $n - 1$ .

There is a useful generalization of (2.14) that we shall now prove.

**Theorem 2.2** *We have the following formula for the determinant of a square matrix  $\overline{A}$  of order  $n + m$  for which every element in the intersection of the first  $n$  columns and last  $m$  rows is zero:*

$$|\overline{A}| = \begin{vmatrix} a_{11} & \cdots & a_{1n} & a_{1n+1} & \cdots & a_{1n+m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} & a_{nn+1} & \cdots & a_{nn+m} \\ 0 & \cdots & 0 & b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & b_{m1} & \cdots & b_{mm} \end{vmatrix}$$

$$= \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} \cdot \begin{vmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mm} \end{vmatrix}. \quad (2.15)$$

*Proof* We again make use of the definition of a determinant, namely formula (2.13), now of order  $n + m$ , and we again employ induction on  $n$ . In our case, the last  $m$  terms of (2.13) are equal to zero, and so we obtain

$$|\bar{A}| = a_{11}\bar{M}_{11} - a_{21}\bar{M}_{21} + a_{31}\bar{M}_{31} - \cdots + (-1)^{n+1}a_{n1}\bar{M}_{n1}. \quad (2.16)$$

It is now clear that  $\bar{M}_{i1}$  is a determinant of the same type as  $\bar{A}$ , but of order  $n - 1 + m$ . Therefore, by the induction hypothesis, we can apply the theorem to this determinant, obtaining

$$|\bar{M}_{i1}| = M_{i1} \cdot \begin{vmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mm} \end{vmatrix}, \quad (2.17)$$

where  $M_{i1}$  has the same meaning as in (2.13) for the determinant  $|A|$ . Substituting expressions (2.17) into (2.16) and using (2.13) for  $|A|$ , we obtain relation (2.15). The theorem is proved.  $\square$

*Remark 2.3* One may well ask why in our definition the first column played a special role and what sort of expressions we might obtain were we to formulate the definition in terms not of the first column, but of the second, third,  $\dots$ , column. As we shall see, the expression obtained will differ from the determinant by at most a sign.

Now let us consider some of the basic properties of determinants. Later on, we shall see that in the theory of determinants, just as in the theory of systems of linear equations, an important role is played by elementary row operations. Let us note that elementary operations like those of type I and type II can be applied to the rows of a matrix whether or not it is the matrix of a system of equations. Theorem 1.15 shows that an *arbitrary* matrix can be transformed into echelon and triangular form.

Therefore, it will be useful to figure out how elementary operations on the rows of a matrix affect the matrix's determinant. In connection with this, we shall introduce some special notation for the rows of a matrix  $A$ : We shall denote by  $\mathbf{a}_i$  the  $i$ th row of  $A$ ,  $i = 1, \dots, n$ . Thus

$$\mathbf{a}_i = (a_{i1}, a_{i2}, \dots, a_{in}).$$

We shall prove several important properties of determinants. We shall prove Properties 2.4, 2.6, and 2.7 below by induction on the order  $n$  of the determinant. For  $n = 1$  (or for Property 2.6, for  $n = 2$ ), these properties are obvious, and we shall omit a proof. We can therefore assume in the proof that the properties have been proved for determinants of order  $n - 1$ .

By definition (2.13), a determinant is a function that assigns to the matrix  $A$  a certain number  $|A|$ . We shall now assume that all the rows of the matrix  $A$  except for one, let us say the  $i$ th, are fixed, and we shall explain how the determinant depends on the elements of the  $i$ th row  $\mathbf{a}_i$ .

*Property 2.4* The determinant of a matrix is a linear function of the elements of an arbitrary row of the matrix.

*Proof* Let us suppose that we wish to prove this property for the  $i$ th row of matrix  $A$ . We shall use formula (2.13) and show that every term in it is a linear function of the elements of the  $i$ th row. For this, it suffices to choose numbers  $d_{1j}, d_{2j}, \dots, d_{nj}$  such that

$$\pm a_{j1} M_{j1} = d_{1j} a_{i1} + d_{2j} a_{i2} + \dots + d_{nj} a_{in}$$

for all  $j = 1, 2, \dots, n$  (see the definition of linear function on p. 2). We begin with the term  $\pm a_{i1} M_{i1}$ . Since the minor  $M_{i1}$  does not depend on the elements of the  $i$ th row—the  $i$ th row is ignored in the calculation—it is simply a constant as a function of the  $i$ th row. Let us set  $d_{1i} = \pm M_{i1}$  and  $d_{2i} = d_{3i} = \dots = d_{ni} = 0$ . Then the first term is represented in the required form, and indeed is a linear function of the  $i$ th row of the matrix  $A$ . For the term  $\pm a_{j1} M_{j1}$ , for  $j \neq i$ , the element  $a_{j1}$  does not appear in the  $i$ th row, but all the elements of the  $i$ th row of matrix  $A$  other than  $a_{i1}$  appear in some row of the minor  $M_{j1}$ . Therefore, by the induction hypothesis,  $M_{j1}$  is a linear function of these elements, that is,

$$M_{j1} = d'_{2j} a_{i2} + \dots + d'_{nj} a_{in}$$

for some numbers  $d'_{2j}, \dots, d'_{nj}$ . Setting  $d_{2j} = a_{j1} d'_{2j}, \dots, d_{nj} = a_{j1} d'_{nj}$ , and  $d_{1j} = 0$ , we convince ourselves that  $a_{j1} M_{j1}$  is a linear function of the  $i$ th row of matrix  $A$ , but this means that such is also the case for the function  $\pm a_{j1} M_{j1}$ . Therefore,  $|A|$  is the sum of linear functions of the elements of the  $i$ th row, and it follows that  $|A|$  is itself a linear function (see p. 4).  $\square$

**Corollary 2.5** *If we apply Theorem 1.3 to a determinant as a function of its  $i$ th row,<sup>1</sup> then we obtain the following:*

1. *Multiplication of each of the elements of the  $i$ th row of a matrix  $A$  by the number  $p$  multiplies the determinant  $|A|$  by the same number.*
2. *If all elements of the  $i$ th row of matrix  $A$  are of the form  $a_{ij} = b_j + c_j$ , then its determinant  $|A|$  is equal to the sum of the determinants of two matrices, in each of which all the elements other than the elements in the  $i$ th row are the same as in the original, and in the  $i$ th row of the first determinant, instead of the elements*

---

<sup>1</sup>We are being a bit sloppy with language here. We have defined the determinant as a function that assigns a number to a matrix, so when we speak of the “rows of a determinant,” this is shorthand for the rows of the underlying matrix.

$a_{ij}$ , one has the numbers  $b_j$ , while in the  $i$ th row of the other one, the numbers are  $c_j$ .

**Property 2.6** The transposition of two rows of a determinant changes its sign.

*Proof* We again begin with formula (2.13). Let us assume that we have interchanged the positions of rows  $j$  and  $j + 1$ . We first consider the term  $a_{i1}M_{i1}$ , where  $i \neq j$  and  $i \neq j + 1$ . Then interchanging the  $j$ th and  $(j + 1)$ st rows does not affect the elements  $a_{i1}$ . As for the minor  $M_{i1}$ , it contains the elements of both the  $j$ th and  $(j + 1)$ st rows of the original matrix (other than the first element of each row), where they again fill two neighboring rows. Therefore, by the induction hypothesis, the minor  $M_{i1}$  changes sign when the rows are transposed. Thus every term  $a_{i1}M_{i1}$  with  $i \neq j$  and  $i \neq j + 1$  changes sign with a transposition of the  $j$ th and  $(j + 1)$ st rows. The remaining terms have the form

$$\begin{aligned} & (-1)^{j+1}a_{j1}M_{j1} + (-1)^{j+2}a_{j+11}M_{j+11} \\ &= (-1)^{j+1}(a_{j1}M_{j1} - a_{j+11}M_{j+11}). \end{aligned} \quad (2.18)$$

With a transposition of the  $j$ th and  $(j + 1)$ st rows, it is easily seen that the terms  $a_{j1}M_{j1}$  and  $a_{j+11}M_{j+11}$  exchange places, which means that the entire expression (2.18) changes sign. This proves Property 2.6.  $\square$

In what follows, a prominent role will be played by the square matrices

$$E = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}, \quad (2.19)$$

all of whose elements on the main diagonal are equal to 1 and all of whose nondiagonal elements are equal to zero. Such a matrix  $E$  is called an *identity* matrix. Of course, for every natural number  $n$  there exists an identity matrix of order  $n$ , and when we wish to emphasize the order of the identity matrix under consideration, we shall write  $E_n$ .

**Property 2.7** The determinant of the identity matrix  $E_n$ , for all  $n \geq 1$ , is equal to 1.

*Proof* In formula (2.13),  $a_{i1} = 0$  if  $i \neq 1$ , and  $a_{11} = 1$ . Therefore,  $|E| = M_{11}$ . The determinant  $M_{11}$  has the same structure as  $|E|$ , but its order is  $n - 1$ . By the induction hypothesis, we may assume that  $M_{11} = 1$ , which means that  $|E| = 1$ .  $\square$

In proving Properties 2.4, 2.6, and 2.7, it was necessary to use definition (2.13). Now we shall prove a series of properties of the determinant that can be formally derived from these first three properties.

*Property 2.8* If all the elements of a row of a matrix are equal to 0, then the determinant of the matrix is equal to 0.

*Proof* Let  $a_{i1} = a_{i2} = \cdots = a_{in} = 0$ . We may set  $a_{ik} = pb_{ik}$ , where  $p = 0$ ,  $b_{ik} \neq 0$ ,  $k = 1, \dots, n$ , and apply the first assertion of Corollary 2.5. We obtain that  $|A| = p|A'|$ , where  $|A'|$  is some other determinant and the number  $p$  is equal to zero. We conclude that  $|A| = 0$ .  $\square$

*Property 2.9* If we transpose any two (not necessarily adjacent) rows of a determinant, then the determinant changes sign.

*Proof* Let us transpose the  $i$ th and  $j$ th rows, where  $i < j$ . The same result can be achieved by successively transposing adjacent rows. Namely, we begin by transposing the  $i$ th and  $(i + 1)$ st rows, then the  $(i + 1)$ st and  $(i + 2)$ nd, and so on until the  $i$ th row has been moved adjacent to the  $j$ th row, that is, into the  $(j - 1)$ st position. At this point, we have carried out  $j - i - 1$  transpositions of adjacent rows. Then we transpose the  $(j - 1)$ st and  $j$ th rows, thereby increasing the number of transpositions to  $j - i$ . We then transpose the  $j$ th row with its successive neighbors so that it occupies the  $i$ th position. In the end, we will have exchanged the positions of the  $i$ th and  $j$ th rows, with all other rows occupying their original positions. In carrying out this process, we have transposed adjacent rows  $(i - j - 1) + 1 + (i - j - 1) = 2(i - j - 1) + 1$  times. This is an odd number. Therefore, by Property 2.6, which asserts that interchanging two rows of a matrix results in a change of sign in the determinant, the result of all transpositions in this process is a change in the determinant's sign.  $\square$

Property 2.9 can also be stated thus: An elementary operation of type I on the rows of a determinant changes its sign.

*Property 2.10* If two rows of a matrix  $A$  are equal, then the determinant  $|A|$  is equal to zero.

*Proof* Let us transpose the two equal rows of  $A$ . Then obviously, the determinant  $|A|$  does not change. But by Property 2.9, the determinant changes sign. But then we have  $|A| = -|A|$ , that is,  $2|A| = 0$ , from which we may conclude that  $|A| = 0$ .  $\square$

*Property 2.11* If an elementary operation of type II is performed on a determinant, it is unchanged.

*Proof* Suppose that after adding  $c$  times the  $j$ th row of  $A$  to the  $i$ th row, we have the determinant  $A'$ . Its  $i$ th row is the sum of two rows, and by the second assertion of Corollary 2.5, we have the equality  $|A'| = D_1 + D_2$ , where  $D_1 = |A|$ . As for the determinant  $D_2$ , it differs from  $|A|$  in that in the  $i$ th row, it has  $c$  times the  $j$ th row. The factor  $c$  can be taken outside the determinant by the first assertion of Corollary 2.5. Then we have a determinant whose  $i$ th and  $j$ th rows are equal.

But by Property 2.10, such a determinant is equal to zero. Hence  $D_2 = 0$ , and so  $|A'| = |A|$ .  $\square$

We remark that the properties proven above give us a very simple method for computing a determinant of order  $n$ . We have only to apply elementary operations to bring the matrix  $A$  into upper triangular form:

$$\bar{A} = \begin{pmatrix} \bar{a}_{11} & \bar{a}_{12} & \cdots & \bar{a}_{1n} \\ 0 & \bar{a}_{22} & \cdots & \bar{a}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \bar{a}_{nn} \end{pmatrix}.$$

Let us suppose that in the process of doing this, we have completed  $t$  elementary operations of type I and some number of operations of type II. Since operations of type II do not change the determinant, and an operation of type I multiplies the determinant by  $-1$ , we have  $|\bar{A}| = (-1)^t |A|$ . We shall now show that

$$|\bar{A}| = \bar{a}_{11} \bar{a}_{22} \cdots \bar{a}_{nn}. \quad (2.20)$$

Then

$$|A| = (-1)^t \bar{a}_{11} \bar{a}_{22} \cdots \bar{a}_{nn}. \quad (2.21)$$

This is a formula for calculating  $|A|$ .

We shall prove formula (2.20) by induction on  $n$ . Since in the matrix  $\bar{A}$ , all elements of the first column except  $\bar{a}_{11}$  are equal to zero, it follows by formula (2.14) that we have the equality

$$|\bar{A}| = \bar{a}_{11} |\bar{A}'|, \quad (2.22)$$

in which the determinant

$$|\bar{A}'| = \begin{vmatrix} \bar{a}_{22} & \bar{a}_{23} & \cdots & \bar{a}_{2n} \\ 0 & \bar{a}_{33} & \cdots & \bar{a}_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \bar{a}_{nn} \end{vmatrix}$$

has a structure analogous to that of the determinant  $|\bar{A}|$ . By the induction hypothesis, we obtain the equality  $|\bar{A}'| = \bar{a}_{22} \bar{a}_{33} \cdots \bar{a}_{nn}$ . Substituting this expression into (2.22) yields the formula (2.20) for  $|A|$ .

The properties of determinants that we have proved allow us to conclude an important theorem on linear equations.

**Theorem 2.12** *A system of  $n$  equations in  $n$  unknowns has a unique solution if and only if the determinant of the matrix of the system is different from zero.*





**Theorem 2.15** *Let  $F(A)$  be a function that assigns to a square matrix  $A$  of order  $n$  a certain number. If this function satisfies properties 1 and 2 above, then there exists a number  $k$  such that*

$$F(A) = k|A|. \quad (2.24)$$

*In this case, the number  $k$  is equal to  $F(E)$ , where  $E$  is the identity matrix.*

*Proof* First of all, we observe that from properties 1 and 2 it follows that the function  $F(A)$  is unchanged if we apply to the matrix  $A$  an elementary operation of type II, and that it changes sign if we apply an elementary operation of type I. This proves that from properties 1 and 2 above, we have the corresponding properties of the determinant (Properties 2.9 and 2.11 of Sect. 2.2).

Let us now bring matrix  $A$  into echelon form using elementary operations. We write the matrix thus obtained in the form

$$\bar{A} = \begin{pmatrix} \bar{a}_{11} & \bar{a}_{12} & \cdots & \bar{a}_{1n} \\ 0 & \bar{a}_{22} & \cdots & \bar{a}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \bar{a}_{nn} \end{pmatrix}, \quad (2.25)$$

whereby we do not, however, assert that  $\bar{a}_{11} \neq 0, \dots, \bar{a}_{nn} \neq 0$ . Such a form can always be obtained, since for a square matrix in echelon form, all elements  $a_{ij}$ ,  $i > j$ , that is, those below the main diagonal, are equal to zero. Let us assume that in the transition from  $A$  to  $\bar{A}$ , we have performed  $t$  elementary operations of type I, while all the other operations were of type II. Since under an elementary operation of type II neither  $F(A)$  nor  $|A|$  is changed, and under elementary operations of type I, both expressions change sign, it follows that

$$|A| = (-1)^t |\bar{A}|, \quad F(A) = (-1)^t F(\bar{A}). \quad (2.26)$$

In order to prove formula (2.24) in the general case, it now suffices to prove it for matrices  $\bar{A}$  of the form (2.25), that is, to establish the equality  $F(\bar{A}) = k|\bar{A}|$ , which, in turn, clearly follows from the relationships

$$|\bar{A}| = \bar{a}_{11}\bar{a}_{22}\cdots\bar{a}_{nn}, \quad F(\bar{A}) = F(E) \cdot \bar{a}_{11}\bar{a}_{22}\cdots\bar{a}_{nn}. \quad (2.27)$$

We observe that the first of these equalities is precisely the equality (2.20) from the previous section. Moreover, it is a consequence of the second equality, since the determinant  $|A|$ , as we have shown, is also a function of type  $F(A)$ , possessing properties 1 and 2. And therefore, having proved the second equality in (2.27) for an arbitrary function  $F(A)$  possessing the given properties, we shall prove this again for the determinant.

It thus remains only to prove the second equality of (2.27). In view of property 1, we can take out from  $F(\bar{A})$  the factor  $\bar{a}_{nn}$ :

$$F(\bar{A}) = \bar{a}_{nn} \cdot F \left( \begin{pmatrix} \bar{a}_{11} & \bar{a}_{12} & \cdots & \bar{a}_{1n} \\ 0 & \bar{a}_{22} & \cdots & \bar{a}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \right).$$

Let us now add to rows  $1, 2, \dots, n-1$  the last row multiplied by the numbers  $-\bar{a}_{1n}, -\bar{a}_{2n}, \dots, -\bar{a}_{n-1n}$  respectively. In this case, all elements, except the elements of the last column, are unchanged, and all the elements of the last column become equal to zero, with the exception of the  $n$ th, which remains equal to 1. Then let us apply analogous transformations to the matrix of smaller size with elements located in the first  $n-1$  rows and columns, and so on. Each time, the number  $\bar{a}_{ii}$  is factored out of  $F$ , and the argument is repeated. After doing this  $n$  times, we obtain

$$F(\bar{A}) = \bar{a}_{nn} \cdots \bar{a}_{11} \cdot F \left( \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \right),$$

which is the second equality of (2.27). □

## 2.4 Expansion of a Determinant Along Its Columns

On the basis of Theorem 2.15, we can answer a question that arose earlier, in Sect. 2.2: does the *first* column play a special role in (2.12) and (2.13) for a determinant of order  $n$ ? To answer this question, let us form an expression analogous to (2.13), but taking instead of the first column, the  $j$ th column. In other words, let us consider the function

$$F(A) = a_{1j}M_{1j} - a_{2j}M_{2j} + \cdots + (-1)^{n+1}a_{nj}M_{nj}. \quad (2.28)$$

It is clear that this function assigns to every matrix  $A$  of order  $n$  a specific number. Let us verify that it satisfies conditions 1 and 2 of the previous section. To this end, we have simply to examine the proofs of the properties from Sect. 2.2 and convince ourselves that we never used the fact that it was precisely the elements of the *first* column that were multiplied by their respective minors. In other words, the proofs of these properties apply word for word to the function  $F(A)$ . By Theorem 2.15, we have  $F(A) = k|A|$ , and we have only to determine the number  $k$  in the formula  $k = F(E)$ .

For the matrix  $E$ , all the elements  $a_{ij}$  are equal to zero whenever  $i \neq j$ , and the elements  $a_{jj}$  are equal to 1. Therefore, formula (2.28) reduces to the equality

$F(a) = \pm M_{jj}$ . Since in formula (2.28) the signs alternate, the term  $a_{jj}M_{jj}$  appears with the sign  $(-1)^{j+1}$ . Clearly,  $M_{jj}$  is the determinant of the identity matrix  $E$  of order  $n - 1$ , and therefore,  $M_{jj} = 1$ . As a result, we obtain that  $k = (-1)^{j+1}$ , which means that

$$a_{1j}M_{1j} - a_{2j}M_{2j} + \cdots + (-1)^{n+1}a_{nj}M_{nj} = (-1)^{j+1}|A|.$$

We now move the coefficient  $(-1)^{j+1}$  to the left-hand side:

$$|A| = (-1)^{j+1}a_{1j}M_{1j} + (-1)^{j+2}a_{2j}M_{2j} + \cdots + (-1)^{j+n}a_{nj}M_{nj}. \quad (2.29)$$

We see that the element  $a_{ij}$  is multiplied by the expression  $(-1)^{i+j}M_{ij}$ , which is called its *cofactor* and denoted by  $A_{ij}$ . We have therefore obtained the following result.

**Theorem 2.16** *The determinant of a matrix  $A$  is equal to the sum of the elements from any of its columns each multiplied by its associated cofactor:*

$$|A| = a_{1j}A_{1j} + a_{2j}A_{2j} + \cdots + a_{nj}A_{nj}. \quad (2.30)$$

In this statement, each column plays an identical role to that played by any other column. For the first column, it becomes the formula that defines the determinant. Formulas (2.29) and (2.30) are called the *expansion of the determinant along the  $j$ th column*.

As an application of Theorem 2.16, we can obtain a whole series of new properties of determinants.

**Theorem 2.17** *Properties 2.4, 2.6, 2.7, 2.8, 2.9, 2.10, 2.11 and all their corollaries hold not only for the rows of a determinant, but for the columns as well.*

*Proof* It follows from formula (2.30) that the determinant is a linear function of the elements of the  $j$ th column,  $j = 1, \dots, n$ . Consequently, Property 2.4 holds for the columns.

We shall prove Property 2.6 by induction on the order  $n$  of the determinant. For  $n = 1$ , the assertion is empty. For  $n = 2$ , it can be checked using formula (2.3). Now let  $n > 2$ , and let us assume that we have transposed columns numbered  $k$  and  $k + 1$ . We make use of formula (2.30) for  $j \neq k, k + 1$ . Then both the  $k$ th and the  $(k + 1)$ st columns enter into every minor  $M_{ij}$  ( $i = 1, \dots, n$ ). By the induction hypothesis, under a transposition of two columns, each minor will change sign, which means that the determinant as a whole changes sign, which proves Property 2.6 for columns. We observe that in Property 2.7, the statement does not discuss rows or columns, and the remaining properties follow formally from the first three. Therefore, all seven properties and their corollaries are valid for the columns of a determinant.  $\square$

In analogy to Theorem 2.15, from Theorem 2.17 it follows that any multilinear antisymmetric function<sup>2</sup> of the columns of a matrix must be proportional to the determinant function of the matrix. Consequently, we have the analogue of formula (2.24), where the function  $F(A)$  satisfies properties 1 and 2, reformulated for columns. In this case, the value  $k$ , as can easily be seen, remains the same. In particular, for an arbitrary index  $i = 1, \dots, n$ , we have the formula, analogous to (2.30),

$$|A| = a_{i1}A_{i1} + a_{i2}A_{i2} + \dots + a_{in}A_{in}. \quad (2.31)$$

It is called the expansion of the determinant  $|A|$  along the  $i$ th row. The formula for the column or row expansion of a determinant has a broad generalization that goes under the name *Laplace's theorem*. It consists in the fact that one has an analogous expansion of a square matrix of order  $n$  not only along a single column (or row), but for an arbitrary number  $m$  of columns,  $1 \leq m \leq n - 1$ . For this, it is necessary only to determine the cofactor not of a single element, but of the minor of arbitrary order  $m$ . Laplace's theorem can be proved, for example, by induction on the number  $m$ , but we shall not do this, but rather put off its precise formulation and proof to Sect. 10.5 (p. 379), where it will be obtained as a special case of even more general concepts and results.

*Example 2.18* In Example 1.20 (p. 15), we proved that the problem of interpolation, that is, the search for a polynomial of degree  $n$  that passes through  $n + 1$  given points, has a unique solution. Theorem 2.12 shows that the determinant of the matrix of the corresponding linear system (1.20) is different from zero. Now we can easily calculate this determinant and once again verify this property.

The determinant of the matrix of system (1.20) for  $r = n + 1$  has the form

$$|A| = \begin{vmatrix} 1 & c_1 & c_1^2 & \dots & c_1^n \\ 1 & c_2 & c_2^2 & \dots & c_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & c_n & c_n^2 & \dots & c_n^n \\ 1 & c_{n+1} & c_{n+1}^2 & \dots & c_{n+1}^n \end{vmatrix}. \quad (2.32)$$

It is called the *Vandermonde determinant* of order  $n + 1$ . We shall show that this determinant is equal to the product of all differences  $c_i - c_j$  for  $i > j$ , that is, that it can be written in the following form:

$$|A| = \prod_{i>j} (c_i - c_j). \quad (2.33)$$

We shall prove (2.33) by induction on the number  $n$ . For  $n = 1$ , the result is obvious:

$$\begin{vmatrix} 1 & c_1 \\ 1 & c_2 \end{vmatrix} = c_2 - c_1.$$

---

<sup>2</sup>For the definition and a discussion of antisymmetric functions, see Sect. 2.6.

For the proof of the general case, we use the fact that the determinant does not change under an elementary operation of type II (Property 2.11 from Sect. 2.2), and moreover, from Theorem 2.17, this property holds for columns as well as for rows. We will make use of this by subtracting the  $n$ th column multiplied by  $c_1$  from the  $(n+1)$ st, then the  $(n-1)$ st multiplied by  $c_1$  from the  $n$ th, and so on, all the way to the second column, from which we subtract the first multiplied by  $c_1$ . By the indicated property, the determinant does not change under these operations, but on the other hand, it assumes the form

$$|\bar{A}| = \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & c_2 - c_1 & c_2(c_2 - c_1) & \cdots & c_2^{n-1}(c_2 - c_1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & c_n - c_1 & c_n(c_n - c_1) & \cdots & c_n^{n-1}(c_n - c_1) \\ 1 & c_{n+1} - c_1 & c_{n+1}(c_{n+1} - c_1) & \cdots & c_{n+1}^{n-1}(c_{n+1} - c_1) \end{vmatrix}.$$

Making use of Theorem 2.17, we apply to the first row of the determinant thus obtained (consisting of a single nonzero element) the analogue of formula (2.12). As a result, we obtain

$$|\bar{A}| = \begin{vmatrix} c_2 - c_1 & c_2(c_2 - c_1) & \cdots & c_2^{n-1}(c_2 - c_1) \\ \vdots & \vdots & \ddots & \vdots \\ c_n - c_1 & c_n(c_n - c_1) & \cdots & c_n^{n-1}(c_n - c_1) \\ c_{n+1} - c_1 & c_{n+1}(c_{n+1} - c_1) & \cdots & c_{n+1}^{n-1}(c_{n+1} - c_1) \end{vmatrix}.$$

To the last determinant let us apply Corollary 2.5 of Sect. 2.2 and remove from each row its common factor. We obtain

$$|A| = |\bar{A}| = (c_2 - c_1) \cdots (c_n - c_1)(c_{n+1} - c_1) \begin{vmatrix} 1 & c_2 & \cdots & c_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & c_n & \cdots & c_n^{n-1} \\ 1 & c_{n+1} & \cdots & c_{n+1}^{n-1} \end{vmatrix}. \quad (2.34)$$

The last determinant is a Vandermonde determinant of order  $n$ , and by the induction hypothesis, we can assume that formula (2.33) holds for it. Putting the expression (2.33) for a Vandermonde determinant of order  $n$  into expression (2.34), we obtain the desired formula (2.33) for a Vandermonde determinant of order  $n+1$ . Since we have assumed that all the numbers  $c_1, \dots, c_{n+1}$  are distinct, the product of the differences  $c_i - c_j$  for  $i > j$  must be different from zero, and we obtain a new proof of the result that polynomial interpolation as described has a unique solution.

## 2.5 Cramer's Rule

We are now going to derive explicit formulas for the solution of a system of  $n$  equations in  $n$  unknowns, formulas for which we have developed the theory of de-

terminants. The matrix  $A$  of this system is a square matrix of order  $n$ , and we shall assume that it is not singular.

**Lemma 2.19** *The sum of the elements  $a_{ij}$  of an arbitrary (here the  $j$ th) column of a determinant each multiplied by the cofactor  $A_{ik}$  corresponding to the elements of any other column (here the  $k$ th) is equal to zero:*

$$a_{1j}A_{1k} + a_{2j}A_{2k} + \cdots + a_{nj}A_{nk} = 0 \quad \text{for } k \neq j.$$

*Proof* We replace the  $k$ th column in our determinant  $|A|$  with its  $j$ th column. As a result, we obtain a determinant  $|A'|$  that by Property 2.10 of Sect. 2.2, reformulated for columns, is equal to zero. On the other hand, let us expand the determinant  $|A'|$  along the  $k$ th column. Since in forming the cofactors of this column, the elements of the  $k$ th column cancel, we obtain the same cofactors  $A_{ik}$  as in our original determinant  $|A|$ . Therefore, we obtain

$$|A'| = a_{1j}A_{1k} + a_{2j}A_{2k} + \cdots + a_{nj}A_{nk} = 0,$$

which is what we wished to show.

**Theorem 2.20** (Cramer's rule) *If the determinant of the matrix of a system of  $n$  equations in  $n$  unknowns is different from zero, then its solution is given by*

$$x_k = \frac{D_k}{D}, \quad k = 1, \dots, n, \quad (2.35)$$

where  $D$  is the determinant of the matrix of the system, and  $D_k$  is obtained from  $D$  by replacing the  $k$ th column of the matrix with the column of constant terms.

*Proof* By Theorem 2.12, we know that there is a unique collection of values for  $x_1, \dots, x_n$  that transforms the system

[illegible]

into the identity. Let us determine the unknown  $x_k$  for a given  $k$ .

To do so, we shall proceed exactly as in the case of systems of two and three equations from Sect. 2.1: we multiply the  $i$ th equation by the cofactor  $A_{ik}$  and then sum all the resulting equations. After this, the coefficient of  $x_k$  will have the form

$$a_{1k}A_{1k} + \cdots + a_{nk}A_{nk} = D.$$

The coefficient of  $x_j$  for  $j \neq k$  will assume the form

$$a_{1j}A_{1k} + \cdots + a_{nj}A_{nk}.$$

By Lemma 2.19, this number is equal to zero. Finally, for the constant term we obtain the expression

$$b_1 A_{1k} + \cdots + b_n A_{nk}.$$

But it is precisely this expression that we obtain if we expand the determinant  $D_k$  along its  $k$ th column. Therefore, we arrive at the equality

$$Dx_k = D_k,$$

and since  $D \neq 0$ , we have  $x_k = D_k/D$ . This is formula (2.35).  $\square$

## 2.6 Permutations, Symmetric and Antisymmetric Functions

A careful study of the properties of determinants leads to a number of important mathematical concepts relating to arbitrary finite sets that in fact could have been presented earlier.

Let us recall that in Sect. 1.1 we studied linear functions as functions of rows of length  $n$ . In Sect. 2.2 we looked at determinants as functions of square matrices. If we are interested in the dependence of the determinant on the rows of its underlying matrix, then it is possible to consider it as a function of its  $n$  rows:  $|A| = F(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ , where for the matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

we denote by  $\mathbf{a}_i$  its  $i$ th row:

$$\mathbf{a}_i = (a_{i1}, a_{i2}, \dots, a_{in}).$$

Here we encounter the notion of a function  $F(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$  of  $n$  elements of a set  $M$  as a rule that assigns to any  $n$  elements from  $M$ , taken in a particular order, some element of another set  $N$ . Thus,  $F$  is a mapping from  $M^n$  to  $N$  (see p. xvii). In our case,  $M$  is the set of all rows of fixed length  $n$ , and  $N$  is the set of all numbers.

Let us introduce some necessary notation for the sequel. Let  $M$  be a finite set consisting of  $n$  elements  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ .

**Definition 2.21** A function on the  $n$  elements of a set  $M$  is said to be *symmetric* if it is unchanged under an arbitrary rearrangement of its arguments.

After numbering the  $n$  elements of the set  $M$  with the indices  $1, 2, \dots, n$ , we can consider that we have arranged them in order of increasing index. A permutation of them can be considered a rearrangement in another order, which we shall write as

follows. Let  $j_1, j_2, \dots, j_n$  represent the same numbers  $1, 2, \dots, n$ , but perhaps listed in a different order. In this case, we shall say that  $(j_1, j_2, \dots, j_n)$  is a *permutation* of the numbers  $(1, 2, \dots, n)$ . Analogously, we shall say that  $(a_{j_1}, a_{j_2}, \dots, a_{j_n})$  is a permutation of the elements  $(a_1, a_2, \dots, a_n)$ .

Thus the definition of a symmetric function can be written as the equality

$$F(a_{j_1}, a_{j_2}, \dots, a_{j_n}) = F(a_1, a_2, \dots, a_n) \quad (2.36)$$

for all permutations  $(j_1, j_2, \dots, j_n)$  of the numbers  $(1, 2, \dots, n)$ .

In order to determine whether one is dealing with a symmetric function, it is not necessary to verify equality (2.36) for all permutations  $(j_1, j_2, \dots, j_n)$ , but instead we can limit ourselves to certain permutations of the simplest form.

**Definition 2.22** A permutation of two elements of the set  $(a_1, a_2, \dots, a_n)$  is called a *transposition*.

A transposition under which the  $i$ th and  $j$ th elements (that is,  $a_i$  and  $a_j$ ) are transposed will be denoted by  $\tau_{i,j}$ . Clearly, we may always assume that  $i < j$ .

We have the following simple fact about permutations.

**Theorem 2.23** *From any arrangement  $(i_1, i_2, \dots, i_n)$  of distinct natural numbers taking values from 1 to  $n$ , it is possible to obtain an arbitrary permutation  $(j_1, j_2, \dots, j_n)$  by carrying out a certain number of transpositions.*

*Proof* We shall use induction on  $n$ . For  $n = 1$ , the assertion of the theorem is a tautology: there exists only one permutation, and so it is unnecessary to introduce any transpositions at all. In the general case ( $n > 1$ ), let us suppose that  $j_1$  stands at the  $k$ th position in the permutation  $(i_1, i_2, \dots, i_n)$ , that is,  $j_1 = i_k$ . We will perform the transposition  $\tau_{1,k}$  on this permutation. If  $j_1 = i_1$ , then it is not necessary to perform any transposition at all. We obtain the permutation  $(j_1, i_2, \dots, i_1, \dots, i_n)$ , where  $j_1$  is in the first position, and  $i_1$  is in the  $k$ th position. Now we need to use transpositions to obtain from the permutation  $(j_1, i_2, \dots, i_1, \dots, i_n)$  the second permutation,  $(j_1, j_2, \dots, j_n)$ , given in the statement of the theorem.

If we cancel  $j_1$  from both permutations, then what remains is a permutation of the numbers  $\alpha$  such that  $1 \leq \alpha \leq n$  and  $\alpha \neq j_1$ . To these two permutations now consisting of only  $n - 1$  numbers, we can apply the induction hypothesis and obtain the second permutation from the first. Beginning with the transposition  $\tau_{1,k}$ , we can thus obtain from the permutation  $(i_1, i_2, \dots, i_n)$  the permutation  $(j_1, j_2, \dots, j_n)$ . In some cases, it will not be necessary to apply a transposition (for example, if  $j_1 = i_1$ ). The limiting case can also be encountered in which it will not be necessary to use any transpositions at all. It is easy to see that such occurs only for  $i_1 = j_1, i_2 = j_2, \dots, i_n = j_n$ . The assertion of the theorem is true in this case, but the set of transpositions used is empty.  $\square$

This very simple argument can be illustrated as follows. Let us suppose that at a concert, the invited guests sit down in the first row, but not in the order indicated on



the administrator's guest list. How can he achieve the requisite ordering? Obviously, he may identify the guest who should be sitting in the first position and ask that person to change seats with the person sitting in the first chair. He will then do likewise with the guests who occupy the second, third, and so on, places, and in the end will have achieved the required order.

It follows from Theorem 2.23 that in determining that a function is symmetric, it suffices to verify equality (2.36) for permutations obtained from the permutation  $(1, 2, \dots, n)$  by a single transposition, that is, to check that

$$F(a_1, \dots, a_i, \dots, a_j, \dots, a_n) = F(a_1, \dots, a_j, \dots, a_i, \dots, a_n)$$

for arbitrary  $a_1, \dots, a_n$ ,  $i$ , and  $j$ . Indeed, if this property is satisfied, then applying various transpositions successively to the argument of the function  $F(a_1, \dots, a_n)$ , we will always obtain the same function, and by Theorem 2.23, we will finally obtain the function  $F(a_{j_1}, \dots, a_{j_n})$ .

For example, for  $n = 3$ , we have three transpositions:  $\tau_{1,2}$ ,  $\tau_{2,3}$ ,  $\tau_{1,3}$ . For the function  $F(a_1, a_2, a_3) = a_1a_2 + a_1a_3 + a_2a_3$ , for example, under the transposition  $\tau_{1,2}$ , the term  $a_1a_2$  remains unchanged, but the other two terms exchange places. The same sort of thing transpires for the other transpositions. Therefore, our function is symmetric.

We now consider a class of functions that in a certain sense are the opposite of symmetric.

**Definition 2.24** A function on  $n$  elements of a set  $M$  is said to be *antisymmetric* if under a transposition of its elements it changes sign.

In other words,

$$F(a_1, \dots, a_i, \dots, a_j, \dots, a_n) = -F(a_1, \dots, a_j, \dots, a_i, \dots, a_n)$$

for any  $a_1, \dots, a_n$ ,  $i$ , and  $j$ .

The notions of symmetric and antisymmetric function play an extremely important role in mathematics and mathematical physics. For example, in quantum mechanics, the state of a certain physical quantity in a system consisting of  $n$  (generally a very large number) elementary particles  $p_1, \dots, p_n$  of a single type is described by a *wave function*  $\psi(p_1, \dots, p_n)$  that depends on these particles and assumes complex values. In a certain sense, in the “general case,” a wave function is symmetric or antisymmetric, and which of these two possibilities is realized depends only on the type of particle: photons, electrons, and so on. If the wave function is symmetric, then the particles are called *bosons*, and in this case, we say that the quantum-mechanical system under consideration is subordinate to the *Bose–Einstein statistics*. On the other hand, if the wave function is antisymmetric, then the particles are called *fermions*, and we say that the system is subordinate to the *Fermi–Dirac statistics*.<sup>3</sup>

---

<sup>3</sup>For example, photons are bosons, and the particles that make up the atom—electrons, protons, and neutrons—are fermions.

We shall return to a consideration of symmetric and antisymmetric functions in the closing chapters of this book. For now, we would like to answer the following question: How is an antisymmetric function transformed under an arbitrary permutation of the indices? In other words, we would like to express  $F(\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_n})$  in terms of  $F(\mathbf{a}_1, \dots, \mathbf{a}_n)$  for an arbitrary permutation  $(i_1, \dots, i_n)$  of the indices  $(1, \dots, n)$ . To answer this, we again turn to Theorem 2.23, according to which the permutation  $(i_1, \dots, i_n)$  can be obtained from the permutation  $(1, \dots, n)$  via a certain number  $k$ , let us say) of transpositions. However, the hallmark of an antisymmetric function is that it changes sign under the transposition of two of its arguments. After  $k$  transpositions, therefore, it will have been altered by the sign  $(-1)^k$ , and we obtain the relationship

$$F(\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_n}) = (-1)^k F(\mathbf{a}_1, \dots, \mathbf{a}_n), \quad (2.37)$$

where the collection of elements  $\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_n}$  from the set  $M$  is obtained from the collection  $\mathbf{a}_1, \dots, \mathbf{a}_n$  by means of the permutation under consideration consisting of  $k$  transpositions.

The relationship (2.37) has about it a certain ambiguity. Namely, the number  $k$  indicates the number of transpositions that are executed in passing from  $(1, \dots, n)$  to the permutation  $(i_1, \dots, i_n)$ . But such a passage can in general be accomplished in a variety of ways, and so the required number  $k$  of transpositions can assume a number of different values. For example, to pass from  $(1, 2, 3)$  to the permutation  $(3, 2, 1)$ , we could begin with the transposition  $\tau_{1,2}$ , obtaining  $(2, 1, 3)$ . Then we could apply the transposition  $\tau_{2,3}$  and arrive at the permutation  $(2, 3, 1)$ . And finally, again carrying out the transposition  $\tau_{1,2}$ , we would arrive at the permutation  $(3, 2, 1)$ . Altogether, we carried out three transpositions. On the other hand, we can carry out a single transposition  $(\tau_{1,3})$ , which from  $(1, 2, 3)$  gives us immediately the permutation  $(3, 2, 1)$ . Nevertheless, let us note that we have not produced any inconsistency with (2.37), since both values of  $k$ , namely 3 and 1, are odd, and therefore in both cases, the coefficient  $(-1)^k$  has the same value.

Let us show that the parity of the number of transpositions used in passing from one given permutation to another depends only on the permutations themselves and not on the choice of transpositions. Let us suppose that we have an antisymmetric function  $F(\mathbf{a}_1, \dots, \mathbf{a}_n)$  that depends on  $n$  elements of a set  $M$  and is not identically zero. This last assumption means that there exists a set of distinct elements  $\mathbf{a}_1, \dots, \mathbf{a}_n$  from the set  $M$  such that  $F(\mathbf{a}_1, \dots, \mathbf{a}_n) \neq 0$ . On applying the permutation  $(i_1, \dots, i_n)$  to this set of elements, we obtain  $(\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_n})$ , with the values  $F(\mathbf{a}_1, \dots, \mathbf{a}_n)$  and  $F(\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_n})$  related by (2.37). If we can obtain the permutation  $(i_1, \dots, i_n)$  from  $(1, \dots, n)$  in two different ways, that is, using  $k$  and  $l$  transpositions, then from formula (2.37) we have the equality  $(-1)^k = (-1)^l$ , since  $F(\mathbf{a}_1, \dots, \mathbf{a}_n) \neq 0$ , and therefore the numbers  $k$  and  $l$  have the same parity, that is, either both are even or both are odd.

But there is a function known to us that possesses this property, namely the determinant (as a function of the rows of a matrix)! Indeed, Property 2.9 from Sect. 2.2 asserts that the determinant is an antisymmetric function of its rows. This function is nonzero for some  $\mathbf{a}_1, \dots, \mathbf{a}_n$ . For example,  $|E| = 1$ . In other words, to prove our

assertion, it suffices to consider the determinant of the matrix  $E$  as an antisymmetric function of its  $n$  rows  $\mathbf{e}_i = (0, \dots, 1, \dots, 0)$ , where there is a 1 in the  $i$ th place and zeros in the other places, for  $i = 1, \dots, n$ . (In the course of our argument, these rows will be transposed, so that in fact, we shall consider determinants of matrices more complex than  $E$ .) Thus by a rather roundabout route, using properties of the determinant, we have obtained the following property of permutations.

**Theorem 2.25** *For any passage from the permutation  $(1, \dots, n)$  to the permutation  $\mathbf{J} = (j_1, \dots, j_n)$  by means of transpositions (which is always possible, thanks to Theorem 2.23), the parity of the number of transpositions will be the same as for any other passage between these two permutations.*

Thus the set of all permutations of  $n$  items can be divided into two classes: those that can be obtained from the permutation  $(1, \dots, n)$  by means of an even number of transpositions and those that can be obtained with an odd number of transpositions. Permutations of the first type are called *even*, and those of the second type are called *odd*. If some permutation  $\mathbf{J}$  is obtained by  $k$  transpositions, then we introduce the notation

$$\varepsilon(\mathbf{J}) = (-1)^k.$$

In other words, for an even permutation  $\mathbf{J}$ , the number  $\varepsilon(\mathbf{J})$  is equal to 1, and for an odd permutation, we have  $\varepsilon(\mathbf{J}) = -1$ .

We have proved the consistency of the notion of even and odd permutation in a rather roundabout way, using the properties of the determinant. In fact, it would have sufficed for us to produce any antisymmetric function not identically zero, and we used one that was familiar to us: the determinant as a function of its rows. We could have invoked a simpler function. Let  $M$  be a set of numbers, and for  $x_1, \dots, x_n \in M$ , we set

$$\begin{aligned} F(x_1, \dots, x_n) &= (x_2 - x_1)(x_3 - x_1) \cdots (x_n - x_1) \cdots (x_n - x_{n-1}) \\ &= \prod_{i>j} (x_i - x_j). \end{aligned} \tag{2.38}$$

Let us verify that this function is antisymmetric. To this end, we introduce the following lemma.

**Lemma 2.26** *Any transposition can be obtained as the result of an odd number of transpositions of adjacent elements, that is, transpositions of the form  $\tau_{k,k+1}$ .*

We actually proved this statement in essence in Sect. 2.2 when we derived Property 2.9 from Property 2.6. There we did not use the term “transposition,” and instead we spoke about interchanging the rows of a determinant. But that very simple proof can be applied to the elements of any set, and therefore we shall not repeat the argument.

Thus it suffices to prove that the function (2.38) changes sign under the exchange of  $x_k$  and  $x_{k+1}$ . But in this case, the factors  $(x_i - x_j)$  for  $i \neq k, k+1, j \neq k, k+1$ , on the right-hand side of the equation do not change at all. The factors  $(x_i - x_k)$  and  $(x_i - x_{k+1})$  for  $i > k+1$  change places, as do  $(x_k - x_j)$  and  $(x_{k+1} - x_j)$  for  $j < k+1$  also. There remains a single factor  $(x_{k+1} - x_k)$ , which changes sign. It is also clear that the function (2.38) differs from zero for any distinct set of values  $x_1, \dots, x_n$ .

We can now apply formula (2.37) to the function given by relation (2.38), by which we proved Theorem 2.25, which means that the notion of the parity of a permutation is well defined. We note, however, that our “simpler” method is very close to our “roundabout” way with which we began, since formula (2.38) defines the Vandermonde determinant of order  $n$  (see formula (2.33) in Sect. 2.4). Let us choose the numbers  $x_i$  in such a way that  $x_1 < x_2 < \dots < x_n$  (for example, we may set  $x_i = i$ ). Then on the right-hand side of relation (2.38), all factors will be positive.

Let us now write down the analogous relation for  $F(x_{i_1}, \dots, x_{i_n})$ . Since the permutation  $(i_1, \dots, i_n)$  assigns the number  $x_{i_k}$  to the number  $x_k$ , from (2.37), we obtain

$$F(x_{i_1}, \dots, x_{i_n}) = \prod_{k>l} (x_{i_k} - x_{i_l}). \quad (2.39)$$

The sign of  $F(x_{i_1}, \dots, x_{i_n})$  is determined by the number of negative factors on the right-hand side of (2.39). Indeed,  $F(x_{i_1}, \dots, x_{i_n}) > 0$  if the number of factors is even, while  $F(x_{i_1}, \dots, x_{i_n}) < 0$  if it is odd. Negative factors  $(x_{i_k} - x_{i_l})$  arise whenever  $x_{i_k} < x_{i_l}$ , and in view of the choice  $x_1 < x_2 < \dots < x_n$ , this means that  $i_k < i_l$ . It follows that to the negative factors  $(x_{i_k} - x_{i_l})$  there correspond those pairs of numbers  $k$  and  $l$  for which  $k > l$  and  $i_k < i_l$ . In this case, we say that the numbers  $i_k$  and  $i_l$  in the permutation  $(i_1, \dots, i_n)$  stand in *reverse order*, or that they form an *inversion*. Thus a permutation is even or odd according to whether it contains an even or odd number of inversions. For example, in the permutation  $(4, 3, 2, 5, 1)$ , the inversions are the pairs  $(4, 3)$ ,  $(4, 2)$ ,  $(4, 1)$ ,  $(3, 2)$ ,  $(3, 1)$ ,  $(2, 1)$ ,  $(5, 1)$ . In all, there are seven of them, which means that  $F(4, 3, 2, 5, 1) < 0$ , and the permutation  $(4, 3, 2, 5, 1)$  is odd.

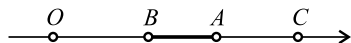
Using these concepts, we can now formulate the following theorem.

**Theorem 2.27** *The determinant of a square matrix of order  $n$  is the unique function  $F(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$  of  $n$  rows of length  $n$  that satisfies the following conditions:*

- (a) *It is linear as a function of an arbitrary row.*
- (b) *It is antisymmetric.*
- (c)  *$F(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) = 1$ , where  $\mathbf{e}_i$  is the row with 1 in the  $i$ th place and zeros in all other places.*

This is the most “scientific,” though far from the simplest, definition of the determinant.

In this section, we have not presented a single new property of the determinant, instead discussing in detail its property of being an antisymmetric function of its

**Fig. 2.2** Path length

rows. The reason for this is that the property of antisymmetry of the determinant is connected with a large number of questions in mathematics. For example, in Sect. 2.1, we introduced determinants of orders 2 and 3. They have an important geometric significance, expressing the area and volume of simple geometric figures (Figs. 2.1(a) and (b)).

But here we encounter a paradoxical situation: Sometimes, one obtains for the area (or volume) a negative value. It is easy to see that we obtain a positive or negative value for the area of triangle  $OAB$  (or the volume of the tetrahedron  $OABC$ ) depending on the order of the vertices  $A, B$  (or  $A, B, C$ ). More precisely, the area of triangle  $OAB$  is positive if we can obtain the ray  $OA$  from  $OB$  by rotating it clockwise through the triangle, while the area is negative if we obtain  $OA$  by rotating  $OB$  counterclockwise through the triangle (in other words, the rotation is always through an angle of measure less than  $\pi$ ). Thus the determinant expresses the area of a triangle (with coefficient  $\frac{1}{2}$ ) with a given ordering of the sides, and the area changes sign if we reverse the order. That is, it is an antisymmetric function.

In the case of volume, choosing the order of the vertices is connected to the concept of *orientation* of space. The same concept appears as well in hyperspaces of dimension  $n > 3$ , but for now, we shall not go too deeply into such questions; we shall return to them in Sects. 4.4 and 7.3. Let us say only that this concept is necessary for constructing the theory of volumes and the theory of integration. In fact, the notion of orientation arises already in the case  $n = 1$ , when we consider the length of an interval  $OA$  (where  $O$  is the origin of the line, namely the point 0, and the point  $A$  has the coordinate  $x$ ) to be the determinant  $x$  of order 1, which will be positive precisely when  $A$  lies to the right of  $O$ . Analogously, if the point  $B$  has coordinate  $y$ , then the length of the segment  $AB$  is equal to  $y - x$ , which will be positive only if  $B$  lies to the right of  $A$ . Thus the length of a segment depends on the ordering of its endpoints, and it changes sign if the endpoints exchange places (thus length is an antisymmetric function). It is only by a similar convention that we can say that the length of  $OABC$  is equal to the length of  $OC$  (Fig. 2.2). And if we were to use only positive lengths, then we would end up with the length of  $OABC$  being given by the expression  $|OA| + |AB| + |BA| + |AC| = |OC| + 2|AB|$ .

## 2.7 Explicit Formula for the Determinant

Formula (2.12), which we used in Sect. 2.2 to compute the determinant of order  $n$ , expresses that determinant in terms of determinants of smaller orders. It is assumed that this method can be applied in turn to these smaller determinants, and passing to determinants of smaller and smaller orders, to arrive at a determinant of order 1, which for the matrix  $(a_{11})$  is equal to  $a_{11}$ . We thereby obtain an expression for the

determinant of the matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

in terms of its elements. This expression is rather complicated, and for deriving the properties of determinants it is simpler to use the inductive procedure given in Sect. 2.2. But now we are ready to discover this complicated definition. First of all, let us prove a lemma, which appears obvious at first glance but nonetheless requires proof (though it is very simple).

**Lemma 2.28** *If the linear function  $f(\mathbf{x})$  for a row  $\mathbf{x}$  of length  $n$  is written in two ways,*

$$f(\mathbf{x}) = \sum_{i=1}^n a_i x_i, \quad f(\mathbf{x}) = \sum_{i=1}^n b_i x_i,$$

*then  $a_1 = b_1, a_2 = b_2, \dots, a_n = b_n$ .*

*Proof* Both of the equations for  $f(\mathbf{x})$  must hold for arbitrary  $\mathbf{x}$ . Let us suppose in particular that  $\mathbf{x} = \mathbf{e}_i = (0, \dots, 1, \dots, 0)$ , where 1 is located in the  $i$ th position (we have already encountered the rows  $\mathbf{e}_i$  in the proof of Theorem 1.3). Then from the initial supposition, we obtain that  $f(\mathbf{e}_i) = a_i$ , and from the second, that  $f(\mathbf{e}_i) = b_i$ . Therefore,  $a_i = b_i$  for all  $i$ , which is what was to be proved.  $\square$

We shall consider the determinant  $|A|$  as a function of the rows  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  of the matrix  $A$ . As shown in Sect. 2.2, the determinant is a linear function of any row of the matrix. A function from any number  $m$  of rows all of length  $n$  is said to be *multilinear* if it is linear in each row (with the other rows held fixed).

**Theorem 2.29** *A multilinear function  $F(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m)$  can be expressed in the form*

$$F(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m) = \sum_{(i_1, i_2, \dots, i_m)} \alpha_{i_1, i_2, \dots, i_m} a_{1i_1} a_{2i_2} \cdots a_{mi_m}, \quad (2.40)$$

*if as usual,  $\mathbf{a}_i = (a_{i1}, a_{i2}, \dots, a_{in})$ , and the sum is taken over arbitrary collections of numbers  $(i_1, i_2, \dots, i_m)$  from the set  $1, 2, \dots, n$ , where  $\alpha_{i_1, i_2, \dots, i_m}$  are certain coefficients that depend only on the function  $F$  and not on the rows  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ .*

*Proof* The proof is by induction on the number  $m$ . For  $m = 1$ , the proof of the theorem is obvious by the definition of a linear function. For  $m > 1$ , we shall use

the fact that

$$F(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m) = \sum_{i=1}^n \varphi_i(\mathbf{a}_2, \dots, \mathbf{a}_m) a_{1i} \quad (2.41)$$

for arbitrary  $\mathbf{a}_1$ , where the coefficients  $\varphi_i$  depend on  $\mathbf{a}_2, \dots, \mathbf{a}_m$ ; that is, they are functions of these numbers.

Let us verify that all the functions  $\varphi_i$  are multilinear. Let us show, for example, linearity with respect to  $\mathbf{a}_2$ . Using the linearity of the function  $F(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m)$  with respect to  $\mathbf{a}_2$ , we obtain

$$F(\mathbf{a}_1, \mathbf{a}'_2 + \mathbf{a}''_2, \dots, \mathbf{a}_m) = F(\mathbf{a}_1, \mathbf{a}'_2, \dots, \mathbf{a}_m) + F(\mathbf{a}_1, \mathbf{a}''_2, \dots, \mathbf{a}_m),$$

or

$$\sum_{i=1}^n \varphi_i(\mathbf{a}'_2 + \mathbf{a}''_2, \dots, \mathbf{a}_m) x_i = \sum_{i=1}^n (\varphi_i(\mathbf{a}'_2, \dots, \mathbf{a}_m) + \varphi_i(\mathbf{a}''_2, \dots, \mathbf{a}_m)) x_i$$

for  $x_i = a_{1i}$ , that is, for arbitrary  $x_i$ . From this, by the lemma, we obtain

$$\varphi_i(\mathbf{a}'_2 + \mathbf{a}''_2, \dots, \mathbf{a}_m) = \varphi_i(\mathbf{a}'_2, \dots, \mathbf{a}_m) + \varphi_i(\mathbf{a}''_2, \dots, \mathbf{a}_m).$$

In precisely the same way, we can verify the second property of linear functions in Theorem 1.3. From this theorem it is seen that the functions  $\varphi_i(\mathbf{a}_2, \dots, \mathbf{a}_m)$  are linear with respect to  $\mathbf{a}_2$ , and analogously that they are multilinear. Now by the induction hypothesis, we have for each of them the expression

$$\varphi_i(\mathbf{a}_2, \dots, \mathbf{a}_m) = \sum_{(i_2, \dots, i_m)} \beta_{i_2, \dots, i_m}^i a_{2i_2} \cdots a_{mi_m} \quad (2.42)$$

(the index  $i$  in  $\beta_{i_2, \dots, i_m}^i$  indicates that these constants are connected with the function  $\varphi_i$ ). To complete the proof, it remains for us, changing notation, to set  $i = i_1$ , to substitute the expressions (2.42) into (2.41), and set  $\beta_{i_2, \dots, i_m}^{i_1} = \alpha_{i_1, i_2, \dots, i_m}$ .  $\square$

*Remark 2.30* The constants  $\alpha_{i_1, i_2, \dots, i_m}$  in the relationship (2.40) can be found from the formulas

$$\alpha_{i_1, i_2, \dots, i_m} = F(\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \dots, \mathbf{e}_{i_m}), \quad (2.43)$$

where  $\mathbf{e}_j$  again denotes the row  $(0, \dots, 1, \dots, 0)$ , in which there is a 1 in the  $j$ th position and zeros everywhere else.

Indeed, if we substitute  $\mathbf{a}_1 = \mathbf{e}_{i_1}$ ,  $\mathbf{a}_2 = \mathbf{e}_{i_2}$ ,  $\dots$ ,  $\mathbf{a}_m = \mathbf{e}_{i_m}$  in the relationship (2.40), then the term  $a_{1i_1} a_{2i_2} \cdots a_{mi_m}$  becomes 1, while the remaining products  $a_{1j_1} a_{2j_2} \cdots a_{mj_m}$  are equal to 0. This proves (2.43).

Let us now apply Theorem 2.29 and (2.43) to the determinant  $|A|$  as a function of the rows  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  of the matrix  $A$ . Since we know that the determinant is a

multilinear function, it must satisfy the relationship (2.40) ( $m = n$ ), and the coefficients  $\alpha_{i_1, i_2, \dots, i_n}$  can be determined from formula (2.43). Consequently,  $\alpha_{i_1, i_2, \dots, i_n}$  is equal to the determinant  $|E_{i_1, i_2, \dots, i_n}|$  of the matrix whose first row is equal to  $\mathbf{e}_{i_1}$ , the second is  $\mathbf{e}_{i_2}$ , ..., and the  $n$ th is  $\mathbf{e}_{i_n}$ . If any of the numbers  $i_1, i_2, \dots, i_n$  are equal, then  $|E_{i_1, i_2, \dots, i_n}| = 0$ , in view of Property 2.10 of Sect. 2.2. It thus remains to examine the determinant  $|E_{i_1, i_2, \dots, i_n}|$  in the case that  $(i_1, i_2, \dots, i_n)$  is a permutation of the numbers  $(1, 2, \dots, n)$ . But this determinant is obtained from the determinant  $|E|$  of the identity matrix if we operate on its rows by the permutation  $(i_1, i_2, \dots, i_n)$ . Furthermore, we know that the determinant is an antisymmetric function of its rows (see Property 2.9 in Sect. 2.2). Therefore, we can apply to it property (2.37) of antisymmetric functions, and we obtain

$$|E_{i_1, i_2, \dots, i_n}| = \varepsilon(\mathbf{I}) \cdot |E|, \quad \text{where } \mathbf{I} = (i_1, i_2, \dots, i_n).$$

Since  $|E| = 1$ , we have the equalities  $\alpha_{i_1, i_2, \dots, i_n} = \varepsilon(\mathbf{I})$  if the permutation  $\mathbf{I}$  is equal to  $(i_1, i_2, \dots, i_n)$ .

As a result, we obtain an expression for the determinant of the matrix  $A$ :

$$|A| = \sum_{\mathbf{I}} \varepsilon(\mathbf{I}) \cdot a_{1i_1} a_{2i_2} \cdots a_{ni_n}, \quad (2.44)$$

where the sum ranges over all permutations  $\mathbf{I} = (i_1, i_2, \dots, i_n)$  of the numbers  $(1, 2, \dots, n)$ . The expression (2.44) is called the *explicit formula for the determinant*. It is worthwhile reformulating this in words:

The determinant of a matrix  $A$  is equal to the sum of terms each of which is the product of  $n$  elements  $a_{ij}$  of the matrix  $A$ , taken one from each row and column. If the factors of such a product are arranged in increasing order of the row numbers, then the term appears with a plus or minus sign depending on whether the corresponding column numbers form an even or odd permutation.

## 2.8 The Rank of a Matrix

In this section, we introduce several fundamental concepts and use them to prove several new results about systems of linear equations.

**Definition 2.31** A matrix whose  $i$ th row coincides with the  $i$ th column of a matrix  $A$  for all  $i$  is called the *transpose* of the matrix  $A$  and is denoted by  $A^*$ .

It is clear that if we denote by  $a_{ij}$  the element located in the  $i$ th row and  $j$ th column of the matrix  $A$ , and by  $b_{ij}$  the corresponding element of the matrix  $A^*$ , then  $b_{ij} = a_{ji}$ . If the matrix  $A$  is of type  $(n, m)$ , then  $A^*$  is of type  $(m, n)$ .

**Theorem 2.32** *The determinant of the transpose of a square matrix is equal to the determinant of the original matrix. That is,  $|A^*| = |A|$ .*



*Proof* Consider the following function of a matrix  $A$ :

$$F(A) = |A^*|.$$

This function exhibits properties 1 and 2 formulated in Sect. 2.3 (page 37). Indeed, the rows of the matrix  $A^*$  are the columns of  $A$ , and thus the assertion that the function  $F(A)$  (that is, the determinant  $|A^*|$  as a function of the matrix  $A$ ) possesses properties 1 and 2 for the rows of the matrix  $A$  is equivalent to the assertion that the determinant  $|A^*|$  possesses the same properties for its columns. This follows from Theorem 2.17. Therefore, Theorem 2.15 is applicable to  $F(A)$ , whence

$$F(A) = k|A|,$$

where  $k = F(E) = |E^*|$ , with  $E$  the  $n \times n$  identity matrix. Clearly,  $E^* = E$ , and therefore,  $k = |E^*| = |E| = 1$ . It follows that  $F(A) = |A|$ , which completes the proof of the theorem.  $\square$

**Definition 2.33** A square matrix  $A$  is said to be *symmetric* if  $A = A^*$ , and *antisymmetric* if  $A = -A^*$ .

It is clear that if  $a_{ij}$  denotes the element located in the  $i$ th row and  $j$ th column of a matrix  $A$ , then the condition  $A = A^*$  can be written in the form  $a_{ij} = a_{ji}$ , while  $A = -A^*$  can be written as  $a_{ij} = -a_{ji}$ . From this last relationship, it follows that all elements  $a_{ii}$  on the main diagonal of an antisymmetric matrix must be equal to zero. Furthermore, it follows from the properties of the determinant that an antisymmetric matrix of odd order is singular. Indeed, if  $A$  is a square matrix of order  $n$ , then from the definition of multiplication of a matrix by a number and the linearity of the determinant in each row, we obtain the relationship  $|-A^*| = (-1)^n |A|$ , from which  $A = -A^*$  yields  $|A| = (-1)^n |A|$ , which in the case of odd  $n$  is possible only if  $|A| = 0$ .

Symmetric and antisymmetric matrices play an important role in mathematics and physics, and we shall encounter them in the following chapters, for example in the study of bilinear forms.

**Definition 2.34** A *minor of order  $r$*  of a matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad (2.45)$$

is a determinant of order  $r$  obtained from the matrix (2.45) by eliminating all entries of the matrix except for those simultaneously in  $r$  given rows and  $r$  given columns. Here we clearly must assume that  $r \leq m$  and  $r \leq n$ .

For example, the minors of order 1 are the individual elements of the matrix, while the unique minor of order  $n$  of a square matrix of order  $n$  is the determinant of the entire matrix.

**Definition 2.35** The *rank* of matrix (2.45) is the maximum over the orders of its nonzero minors.

In other words, the rank is the smallest number  $r$  such that all the minors of rank  $s > r$  are equal to zero or there are no such minors (if  $r = \min\{m, n\}$ ).

Let us note one obvious corollary of Theorem 2.32.

**Theorem 2.36** *The rank of a matrix is not affected by taking the transpose.*

*Proof* The minors of the matrix  $A^*$  are obtained as the transposes of the minors of matrix  $A$  (in taking the transpose, the indices of the rows and columns change places). Therefore, the ranks of the matrices  $A^*$  and  $A$  coincide.  $\square$

Let us recall that in presenting the method of Gaussian elimination in Sect. 1.2, we introduced elementary row operations of types I and II on the equations of a system. These operations changed both the coefficients of the unknowns and the constant terms. If we now focus our attention solely on the coefficients of the unknowns, then we may say that we are carrying out elementary operations on the rows of the matrix of the system. This gives us the possibility of using Gauss's method to determine the rank of a matrix.

A fundamental property of the rank of a matrix is expressed in the following theorem.

**Theorem 2.37** *The rank of a matrix is unchanged under elementary operations on its rows and columns.*

*Proof* We shall carry out the proof for elementary row operations of type II (for type I, the proof is analogous, and even simpler). After adding  $p$  times the  $j$ th row of the matrix  $A$  to the  $i$ th row, we obtain a new matrix; call it  $B$ . We shall denote the rank of a matrix by the operator  $\text{rk}$  and suppose that  $\text{rk } A = r$ . If among the nonzero minors of order  $r$  of the matrix  $A$  there is at least one not containing the  $i$ th row, then it will not be altered by the given operation, and it follows that it will be a nonzero minor of the matrix  $B$ . Therefore, we may conclude that  $\text{rk } B \geq r$ .

Now let us suppose that all nonzero minors of order  $r$  of the matrix  $A$  contain the  $i$ th row. Let  $M$  be one such minor, involving rows numbered  $i_1, \dots, i_r$ , where  $i_k = i$  for some  $k$ ,  $1 \leq k \leq r$ . Let us denote by  $N$  the minor of the matrix  $B$  involving the columns with the same indices as  $M$ . If  $j$  coincides with one of the numbers  $i_1, \dots, i_r$ , then this transformation of the matrix  $A$  is also an elementary transformation of the minor  $M$ , under which it is converted into  $N$ . Since the determinant is unaffected by an elementary transformation of type II, we must have  $N = M$ , whence it follows that  $\text{rk } B \geq r$ .

Now suppose that  $j$  does not coincide with one of the numbers  $i_1, \dots, i_r$ . Let us denote by  $M'$  the minor of the matrix  $A$  involving the same columns as  $M$  and rows numbered  $i_1, \dots, i_{k-1}, j, i_{k+1}, \dots, i_r$ . In other words,  $M'$  is obtained from  $M$  by replacing the  $i_k$ th by the  $j$ th row of the matrix  $A$ . Since the determinant is a linear function of its rows, we therefore have the equality  $N = M + pM'$ . But by our assumption,  $M' = 0$ , since the minor  $M'$  does not contain the  $i$ th row of the matrix  $A$ . Thus we obtain the equality  $N = M$ , from which it follows that  $\text{rk } B \geq r$ .

Thus in all cases we have proved that  $\text{rk } B \geq \text{rk } A$ . However, since the matrix  $A$ , in turn, can be obtained from  $B$  by means of elementary operations of type II, we have the reverse  $\text{rk } A \geq \text{rk } B$ . From this, it clearly follows that  $\text{rk } A = \text{rk } B$ .

By similar arguments, but carried out for operations on the columns, we can show that the rank of a matrix is unchanged under elementary column operations. Furthermore, the assertion for the columns follows from analogous assertions about the rows if we make use of Theorem 2.36.  $\square$

Now we are in a position to formulate answers to the questions that were resolved earlier by Theorems 1.16 and 1.17, without reducing the system to echelon form but instead using explicit expressions that depend on the coefficients. Bringing the system into echelon form will be present in our proofs, but will not appear in the final formulations.

Let us assume that by elementary operations, we have brought a system of equations into echelon form (1.18). By Theorem 2.37, both the rank of the matrix of the system and the rank of the augmented matrix will have remained unchanged. Clearly, the rank of the matrix of (1.18) is equal to  $r$ : a minor at the intersection of the first  $r$  rows and the  $r$  columns numbered  $1, k, \dots, s$  is equal to  $\bar{a}_{11}\bar{a}_{2k} \cdots \bar{a}_{rs}$ , which implies that it is different from zero, and any other minor of greater order must contain a row of zeros and is therefore equal to zero. Therefore, the rank of the matrix of the initial system (1.3) is equal to  $r$ .

The rank of the augmented matrix of system (1.18) is also equal to  $r$  if all the constants  $\bar{b}_{r+1} = \cdots = \bar{b}_n$  are equal to zero or if there are no equations with such numbers ( $m = r$ ). However, if at least one of the numbers  $\bar{b}_{r+1}, \dots, \bar{b}_n$  is different from zero, then the rank of the augmented matrix will be greater than  $r$ . For example, if  $\bar{b}_{r+1} \neq 0$ , then the minor of order  $r + 1$  involving the first  $r + 1$  rows of the augmented matrix and the columns numbered  $1, k, \dots, s, n + 1$  is equal to  $\bar{a}_{11}\bar{a}_{2k} \cdots \bar{a}_{rs}\bar{b}_{r+1}$  and is different from zero. Thus the compatibility criterion formulated in Theorem 1.16 can also be expressed in terms of the rank: the rank of the matrix of system (1.3) must be equal to the rank of the augmented matrix of the system. Since by Theorem 2.37, the rank of the matrix and augmented matrix of the initial system (1.3) are equal to the ranks of the corresponding matrices of (1.18), we obtain the compatibility condition called the *Rouché–Capelli theorem*.

**Theorem 2.38** *The system of linear equations (1.3) is consistent if and only if the rank of the matrix of the system is equal to the rank of the augmented matrix.*

The same considerations make it possible to reformulate Theorem 1.17 in the following form.

**Theorem 2.39** *If the system of linear equations (1.3) is consistent, then it is definite (that is, it has a unique solution) if and only if the rank of the matrix of the system is equal to the number of unknowns.*

We can explain further the significance of the concept of the rank of a matrix in the theory of linear equations by introducing a further notion, one that is important in and of itself.

**Definition 2.40** Suppose we are given  $m$  rows of a given length  $n$ :  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ . A row  $\mathbf{a}$  of the same length is said to be a *linear combination* of  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$  if there exist numbers  $p_1, p_2, \dots, p_m$  such that  $\mathbf{a} = p_1\mathbf{a}_1 + p_2\mathbf{a}_2 + \dots + p_m\mathbf{a}_m$ .

Let us mention two properties of linear combinations.

1. If  $\mathbf{a}$  is a linear combination of the rows  $\mathbf{a}_1, \dots, \mathbf{a}_m$ , each of which, in turn, is a linear combination of the same set of rows  $\mathbf{b}_1, \dots, \mathbf{b}_k$ , then  $\mathbf{a}$  is a linear combination of the rows  $\mathbf{b}_1, \dots, \mathbf{b}_k$ .

Indeed, by the definition of a linear combination, there exist numbers  $q_{ij}$  such that

$$\mathbf{a}_i = q_{i1}\mathbf{b}_1 + q_{i2}\mathbf{b}_2 + \dots + q_{ik}\mathbf{b}_k, \quad i = 1, \dots, m,$$

and numbers  $p_i$  such that  $\mathbf{a} = p_1\mathbf{a}_1 + p_2\mathbf{a}_2 + \dots + p_m\mathbf{a}_m$ . Substituting in the last equality the expression for the rows  $\mathbf{a}_i$  in terms of  $\mathbf{b}_1, \dots, \mathbf{b}_k$ , we obtain

$$\begin{aligned} \mathbf{a} &= p_1(q_{11}\mathbf{b}_1 + q_{12}\mathbf{b}_2 + \dots + q_{1k}\mathbf{b}_k) \\ &\quad + p_2(q_{21}\mathbf{b}_1 + q_{22}\mathbf{b}_2 + \dots + q_{2k}\mathbf{b}_k) + \dots \\ &\quad + p_m(q_{m1}\mathbf{b}_1 + q_{m2}\mathbf{b}_2 + \dots + q_{mk}\mathbf{b}_k). \end{aligned}$$

Removing parentheses and collecting like terms yields

$$\begin{aligned} \mathbf{a} &= (p_1q_{11} + p_2q_{21} + \dots + p_mq_{m1})\mathbf{b}_1 \\ &\quad + (p_1q_{12} + p_2q_{22} + \dots + p_mq_{m2})\mathbf{b}_2 + \dots \\ &\quad + (p_1q_{1k} + p_2q_{2k} + \dots + p_mq_{mk})\mathbf{b}_k, \end{aligned}$$

that is, the expression  $\mathbf{a}$  as a linear combination of the rows  $\mathbf{b}_1, \dots, \mathbf{b}_k$ .

2. When we apply elementary operations to the rows of a matrix, we obtain rows that are linear combinations of the rows of the original matrix.

This is obvious for elementary operations both of type I and of type II.

Let us apply Gaussian elimination to a certain matrix  $A$  of rank  $r$ . Changing the numeration of the rows and columns, we may assume that a nonzero minor of order  $r$  is located in the first  $r$  rows and  $r$  columns of the matrix. Then by elementary

operations on its first  $r$  rows, the matrix is put into the form

$$\bar{A} = \begin{pmatrix} \bar{a}_{11} & \bar{a}_{12} & \cdots & \bar{a}_{1r} & \bar{a}_{1r+1} & \cdots & \bar{a}_{1n} \\ 0 & \bar{a}_{22} & \cdots & \bar{a}_{2r} & \bar{a}_{2r+1} & \cdots & \bar{a}_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \bar{a}_{rr} & \bar{a}_{rr+1} & \cdots & \bar{a}_{rn} \\ \bar{a}_{r+11} & \cdot & \cdots & \cdot & \cdot & \cdots & \bar{a}_{r+1n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \bar{a}_{m1} & \cdot & \cdots & \cdot & \cdot & \cdots & \bar{a}_{mn} \end{pmatrix},$$

where  $\bar{a}_{11} \neq 0, \dots, \bar{a}_{rr} \neq 0$ . We can now subtract from the  $(r+1)$ st row the first row multiplied by a number such that the first element of the row thus obtained is equal to zero, then the second row multiplied by a number such that the second element of the row thus obtained equals zero, and so on, until we obtain the matrix

$$\bar{\bar{A}} = \begin{pmatrix} \bar{a}_{11} & \bar{a}_{12} & \cdots & \bar{a}_{1r} & \bar{a}_{1r+1} & \cdots & \bar{a}_{1n} \\ 0 & \bar{a}_{22} & \cdots & \bar{a}_{2r} & \bar{a}_{2r+1} & \cdots & \bar{a}_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \bar{a}_{rr} & \bar{a}_{rr+1} & \cdots & \bar{a}_{rn} \\ 0 & 0 & \cdots & 0 & \bar{\bar{a}}_{r+1r+1} & \cdots & \bar{\bar{a}}_{r+1n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \bar{\bar{a}}_{mr+1} & \cdots & \bar{\bar{a}}_{mn} \end{pmatrix}.$$

Since the matrix  $\bar{\bar{A}}$  was obtained from  $A$  using a sequence of elementary operations, its rank must be equal to  $r$ .

Let us show that the entire  $(r+1)$ st row of the matrix  $\bar{\bar{A}}$  consists of zeros. Indeed, if there were an element in the row  $\bar{\bar{a}}_{r+1k} \neq 0$  for some  $k = 1, \dots, n$ , then the minor of the matrix  $\bar{\bar{A}}$  formed by the intersection of the first  $r+1$  rows and the columns numbered  $1, 2, \dots, r, k$  would be given by

$$\begin{vmatrix} \bar{a}_{11} & \bar{a}_{12} & \cdots & \bar{a}_{1r} & \bar{a}_{1k} \\ 0 & \bar{a}_{22} & \cdots & \bar{a}_{2r} & \bar{a}_{2k} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \bar{a}_{rr} & \bar{a}_{rk} \\ 0 & 0 & \cdots & 0 & \bar{\bar{a}}_{r+1k} \end{vmatrix} = \bar{a}_{11}\bar{a}_{22}\cdots\bar{a}_{rr}\bar{\bar{a}}_{r+1k} \neq 0,$$

which contradicts the established fact that the rank of  $\bar{\bar{A}}$  is equal to  $r$ .

This result can be formulated thus: If  $\bar{a}_1, \dots, \bar{a}_{r+1}$  are the first  $r+1$  rows of the matrix  $\bar{A}$ , then there exist numbers  $p_1, \dots, p_r$  such that

$$\bar{a}_{r+1} - p_1\bar{a}_1 - \cdots - p_r\bar{a}_r = \mathbf{0}.$$

From this, it follows that  $\bar{a}_{r+1} = p_1\bar{a}_1 + \cdots + p_r\bar{a}_r$ . That is, the row  $\bar{a}_{r+1}$  is a linear combination of the first  $r$  rows of the matrix  $\bar{A}$ . But the matrix  $\bar{A}$  was obtained as the result of elementary operations on the first  $r$  rows of the matrix  $A$ , whence it follows that all rows of the matrices  $\bar{A}$  and  $A$  numbered greater than  $r$  coincide. We see, therefore, that the  $(r+1)$ st row of the matrix  $A$  is a linear combination of the rows  $\bar{a}_1, \dots, \bar{a}_{r+1}$ , each of which, in turn, is a linear combination of the first  $r$  rows of the matrix  $A$ . Consequently, the  $(r+1)$ st row of the matrix  $A$  is a linear combination of its first  $r$  rows.

This line of reasoning carried out for the  $(r+1)$ st row can be applied equally well to any row numbered  $i > r$ . Therefore, every row of the matrix  $A$  is a linear combination of its first  $r$  rows (note that in this case, the *first*  $r$  rows played a special role, since for notational convenience, we numbered the rows and columns in such a way that a nonzero minor was located in the *first*  $r$  rows and first  $r$  columns). In the general case, we obtain the following result.

**Theorem 2.41** *If the rank of a matrix is equal to  $r$ , then all of its rows are linear combinations of some  $r$  rows.*

*Remark 2.42* To put it more precisely, we have shown that if there exists a nonzero minor of order equal to the rank of the matrix, then every row can be written as a linear combination of the rows in which this minor is located.

The application of these ideas to systems of linear equations is based on the following obvious lemma. Here, as in a high-school course, we shall call the equation  $F(\mathbf{x}) = b$  a *corollary* of equations (1.10) if every solution  $\mathbf{c}$  of the system (1.10) satisfies the relationship  $F(\mathbf{c}) = b$ . In other words, this means that if we assign to the system (1.10) one additional equation  $F(\mathbf{x}) = b$ , we obtain an equivalent system.

**Lemma 2.43** *If in the augmented matrix of the system (1.3), some row (say with index  $l$ ) is a linear combination of  $k$  rows, with indices  $i_1, \dots, i_k$ , then the  $l$ th equation of the system is a corollary of the  $k$  equations with those indices.*

*Proof* The proof proceeds by direct verification. To simplify the presentation, let us assume that we are talking about the *first*  $k$  rows of the augmented matrix. Then by definition, there exist  $k$  numbers  $\alpha_1, \dots, \alpha_k$  such that

$$\begin{aligned} & \alpha_1(a_{11}, a_{12}, \dots, a_{1n}, b_1) + \alpha_2(a_{21}, a_{22}, \dots, a_{2n}, b_2) + \cdots \\ & \quad + \alpha_k(a_{k1}, a_{k2}, \dots, a_{kn}, b_k) \\ & = (a_{l1}, a_{l2}, \dots, a_{ln}, b_l). \end{aligned}$$

This means that for every  $i = 1, \dots, n$ , the following equations are satisfied:

$$\begin{cases} \alpha_1 a_{1i} + \alpha_2 a_{2i} + \cdots + \alpha_k a_{ki} = a_{li} & \text{for } i = 1, 2, \dots, n, \\ \alpha_1 b_1 + \alpha_2 b_2 + \cdots + \alpha_k b_k = b_l. \end{cases}$$

Then if we multiply equations numbered  $1, 2, \dots, k$  in our system by the numbers  $\alpha_1, \dots, \alpha_k$  respectively and add the products, we obtain the  $l$ th equation of the system. That is, in the notation of (1.10), we obtain

$$\alpha_1 F_1(\mathbf{x}) + \dots + \alpha_k F_k(\mathbf{x}) = F_l(\mathbf{x}), \quad \alpha_1 b_1 + \dots + \alpha_k b_k = b_l.$$

Substituting here  $\mathbf{x} = \mathbf{c}$ , we obtain that if  $F_1(\mathbf{c}) = b_1, \dots, F_k(\mathbf{c}) = b_k$ , then we have also  $F_l(\mathbf{c}) = b_l$ . That is, the  $l$ th equation is a corollary of the first  $k$  equations.  $\square$

By combining Lemma 2.43 with Theorem 2.41, we obtain the following result.

**Theorem 2.44** *If the rank of the matrix of system (1.3) coincides with the rank of its augmented matrix and is equal to  $r$ , then all the equations of the system are corollaries of some  $r$  equations of the system.*

Therefore, if the rank of the matrix of the combined system (1.3) is equal to  $r$ , then it is equivalent to a system consisting of some  $r$  equations of system (1.3). It is possible to select as these  $r$  equations any such that in the rows with corresponding indices there occurs a nonzero minor of order  $r$  of the matrix of the system (1.3).

## 2.9 Operations on Matrices

In this section, we shall define certain operations on matrices that while simple, are very important for the following presentation. First, we shall define these operations purely formally. Their deeper significance will become clear in the examples presented below, and above all, in the following chapter, where matrices are connected to geometric concepts by linear transformations of vector spaces.

First of all, let us agree that by the equality  $A = B$  for two matrices is meant that  $A$  and  $B$  are matrices of the same type and that their elements (denoted by  $a_{ij}$  and  $b_{ij}$ ) with like indices are equal. That is, if  $A$  and  $B$  each have  $m$  rows and  $n$  columns, then to write  $A = B$  means that the  $m \cdot n$  equalities  $a_{ij} = b_{ij}$  hold for all indices  $i = 1, \dots, m$  and  $j = 1, \dots, n$ .

**Definition 2.45** Let  $A$  be an arbitrary matrix of type  $(m, n)$  with elements  $a_{ij}$ , and let  $p$  be some number. The *product* of the matrix  $A$  and the number  $p$  is the matrix  $B$ , also of type  $(m, n)$ , whose elements satisfy the equations  $b_{ij} = pa_{ij}$ . It is denoted by  $B = pA$ .

Just as is done for numbers, the matrix obtained by multiplying  $A$  by the number  $-1$  is denoted by  $-A$  and is called the *additive inverse* or *opposite*. In the case of the product obtained by multiplying an arbitrary matrix of type  $(m, n)$  by the number  $0$ , we obviously obtain a matrix of the same type, all of whose elements are zero. It is called the *null* or *zero* matrix of type  $(m, n)$  and is denoted by  $0$ .

**Definition 2.46** Let  $A$  and  $B$  be two matrices, each of type  $(m, n)$ , with elements denoted as usual by  $a_{ij}$  and  $b_{ij}$ . The *sum* of  $A$  and  $B$  is the matrix  $C$ , also of type  $(m, n)$ , whose elements  $c_{ij}$  are defined by the formula  $c_{ij} = a_{ij} + b_{ij}$ . This is written as the equality  $C = A + B$ .

Let us emphasize that both sum and equality are defined only for matrices of the same type.

With these definitions in hand, it is now easy to verify that just as in the case of numbers, one has the following rules for removing parentheses:  $(p + q)A = pA + qA$  for any two numbers  $p, q$  and matrices  $A$ , as well as  $p(A + B) = pA + pB$  for any number  $p$  and matrices  $A, B$  of the same type. It is just as easily verified that the addition of matrices does not depend on the order of summation,  $A + B = B + A$ , and that the sum of three (or more) matrices does not depend on the arrangement of parentheses, that is,  $(A + B) + C = A + (B + C)$ . Using addition and multiplication by  $-1$ , it is possible as well to define the *difference* of matrices:  $A - B = A + (-B)$ .

We now define another, the most important of all, operation on matrices, called the *matrix product* or *matrix multiplication*. Like addition, this operation is defined not for matrices of arbitrary type, but only for those whose dimensions obey a certain relationship.

**Definition 2.47** Let  $A$  be a matrix of type  $(m, n)$ , whose elements we shall denote by  $a_{ij}$ , and let  $B$  be a matrix of type  $(n, k)$  with elements  $b_{ij}$  (we observe that here in general, the indices  $i$  and  $j$  of the elements  $a_{ij}$  and  $b_{ij}$  run over different sets of values). The *product* of matrices  $A$  and  $B$  is the matrix  $C$  of type  $(m, k)$  whose elements  $c_{ij}$  are determined by the formula

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}. \quad (2.46)$$

We write the matrix product as  $C = A \cdot B$  or simply  $C = AB$ .

Thus the product of two rectangular matrices  $A$  and  $B$  is defined only in the case that the number of columns of matrix  $A$  is equal to the number of rows of matrix  $B$ , while otherwise, the product is undefined (the reason for this will become clear in the following chapter). The important special case  $n = m = k$  shows that the product of two (and therefore, an arbitrary number of) square matrices of the same order is well defined.

Let us clarify the above definition with the help of some examples.

*Example 2.48* In what follows, we shall frequently encounter matrices of types  $(1, n)$  and  $(n, 1)$ , that is, rows and columns of length  $n$ , often called row vectors and column vectors. For such vectors it is convenient to introduce special notation:

$$\alpha = (\alpha_1, \dots, \alpha_n), \quad [\beta] = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}, \quad (2.47)$$





with coefficients  $b_{ij}$ . Substituting formulas (2.50) into (2.49), we obtain an expression for the variables  $(x_1, \dots, x_m)$  in terms of  $(z_1, \dots, z_k)$ :

$$\begin{aligned} x_i &= a_{i1}(b_{11}z_1 + \dots + b_{1k}z_k) + \dots + a_{in}(b_{n1}z_1 + \dots + b_{nk}z_k) \\ &= (a_{i1}b_{11} + \dots + a_{in}b_{n1})z_1 + \dots + (a_{i1}b_{1k} + \dots + a_{in}b_{nk})z_k. \end{aligned} \quad (2.51)$$

As was done in the previous example, we may write linear substitutions (2.49) and (2.50) in the matrix forms  $[\mathbf{x}] = A[\mathbf{y}]$  and  $[\mathbf{y}] = B[\mathbf{z}]$ , where  $[\mathbf{x}]$ ,  $[\mathbf{y}]$ ,  $[\mathbf{z}]$  are column vectors, whose elements are the corresponding variables, while  $A$  and  $B$  are matrices of types  $(m, n)$  and  $(n, k)$  with elements  $a_{ij}$  and  $b_{ij}$ . Then, by definition (2.46), formula (2.51) assumes the form  $[\mathbf{x}] = C[\mathbf{z}]$ , where the matrix  $C$  is equal to  $AB$ . In other words, *successive application of two linear substitutions gives a linear substitution whose matrix is equal to the product of the matrices of the substitutions.*

**Remark 2.51** All of this makes it possible to formulate a definition of matrix product in terms of linear substitutions: the matrix product of  $A$  and  $B$  is the matrix  $C$  that is the matrix of the substitution obtained by successive applications of two linear substitutions with matrices  $A$  and  $B$ .

This obvious remark makes it possible to give a simple and graphic demonstration of an important property of the matrix product, called *associativity*.

**Theorem 2.52** *Let  $A$  be a matrix of type  $(m, n)$ , and let  $B$  be a matrix of type  $(n, k)$ , and matrix  $D$  of type  $(k, l)$ . Then*

$$(AB)D = A(BD). \quad (2.52)$$

*Proof* Let us first consider the special case  $l = 1$ , that is, the matrix  $D$  in (2.52) is a  $k$ -element column vector. As we have remarked, (2.52) is in this case a simple consequence of the interpretation of the matrix product of  $A$  and  $B$  as the result of carrying out two linear substitutions of the variables; in the notation of Example 2.50, we have simply to substitute  $[\mathbf{z}] = D$  and then use the equalities  $[\mathbf{y}] = B[\mathbf{z}]$ ,  $[\mathbf{x}] = A[\mathbf{y}]$ , and  $[\mathbf{x}] = C[\mathbf{z}]$ .

In the general case, it suffices for the proof of equation (2.52) to observe that the product of matrices  $A$  and  $B$  is reduced to the successive multiplication of the rows of  $A$  by the columns of  $B$ . That is, if we write the matrix  $B$  in column form,  $B = (B_1, \dots, B_k)$ , then  $AB$  can analogously be written in the form  $AB = (AB_1, \dots, AB_k)$ , where each  $AB_i$  is a matrix of type  $(m, 1)$ , that is, also a column vector. After this, the proof of equality (2.52) in the general case is almost self-evident. Let  $D$  consist of  $l$  columns:  $D = (D_1, \dots, D_l)$ . Then on the left-hand side of (2.52), one has the matrix

$$(AB)D = ((AB)D_1, \dots, (AB)D_l),$$

and on the right-hand side, the matrix

$$A(BD) = A(BD_1, \dots, BD_l) = (A(BD_1), \dots, A(BD_l)),$$

and it remains only to use the proved equality (2.52) with  $l = 1$  for each of the column vectors  $D_1, \dots, D_l$ .  $\square$

Let us note that we already considered the associative property in a more abstract form (p. xv). By what was proved there, it follows that the product of any number of factors does not depend on the arrangement of parentheses among them. Thus the associative property makes it possible to compute the product of an arbitrary number of matrices without indicating any arrangement of parentheses (it is necessary only that each pair of associated matrices correspond as to their dimensions so that multiplication is defined). In particular, the result of the product of an arbitrary square matrix by itself an arbitrary number of times is well defined. It is called *exponentiation*.

Just as for numbers, the operations of addition and multiplication of matrices are linked by the relationships

$$A(B + C) = AB + AC, \quad (A + B)C = AC + BC, \quad (2.53)$$

which clearly follow from the definitions. The property (2.53) connecting addition and multiplication is called the *distributive property*.

We mention one important property of multiplication involving the identity matrix: for an arbitrary matrix  $A$  of type  $(m, n)$  and an arbitrary matrix  $B$  of type  $(n, m)$ , the following equalities hold:

$$AE_n = A, \quad E_n B = B.$$

The proofs of both equalities follow from the definition of matrix multiplication, for example, using the rule “row times column.” We see, then, that multiplication by the matrix  $E$  plays the same role as multiplication by 1 among ordinary numbers.

However, another familiar property of multiplication of numbers (called *commutativity*), namely that the product of two numbers is independent of the order in which they are multiplied, is not true for matrix multiplication. This follows at a minimum from the fact that the product  $AB$  of a matrix  $A$  of type  $(n, m)$  and a matrix  $B$  of type  $(l, k)$  is defined only if  $m = l$ . It could well be that  $m = l$  but  $k \neq n$ , and then the matrix product  $BA$  would not be defined, while the product  $AB$  was. But even, for example, in the case  $n = m = k = l = 2$ , with

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad B = \begin{pmatrix} p & q \\ r & s \end{pmatrix},$$

where both products  $AB$  and  $BA$  are defined, we obtain

$$AB = \begin{pmatrix} ap + br & aq + bs \\ cp + dr & cq + ds \end{pmatrix}, \quad BA = \begin{pmatrix} ap + cq & bp + dq \\ ar + cs & br + ds \end{pmatrix},$$

and these are in general unequal matrices. Matrices  $A$  and  $B$  for which  $AB = BA$  are called *commuting matrices*.

In connection with the multiplication of matrices, notation is used that we will introduce only in the special case that we shall actually encounter in what follows. Assume that we are given a square matrix  $A$  of order  $n$  and a natural number  $p < n$ . The elements of the matrix  $A$  located in the first  $p$  rows and first  $p$  columns form a square matrix  $A_{11}$  of order  $p$ . The elements located in the first  $p$  rows and last  $n - p$  columns form a rectangular matrix  $A_{12}$  of type  $(p, n - p)$ . The elements located in the first  $p$  columns and last  $n - p$  rows form a rectangular matrix  $A_{21}$  of type  $(n - p, p)$ . Finally, the elements in the last  $n - p$  rows and last  $n - p$  columns form a rectangular matrix  $A_{22}$  of order  $n - p$ . This can be written as follows:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}. \quad (2.54)$$

Formula (2.54) is called the expression of  $A$  in *block form*, while matrices  $A_{11}$ ,  $A_{12}$ ,  $A_{21}$ ,  $A_{22}$  are the *blocks* of the matrix  $A$ . For example, with these conventions, formula (2.15) takes the form

$$|A| = \begin{vmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{vmatrix} = |A_{11}| \cdot |A_{22}|.$$

Clearly, one can conceive of a matrix  $A$  in block form for a larger number of matrix blocks of various sizes. In addition to the case (2.54) shown above, we shall find ourselves in the situation in which blocks stand on the diagonal:

$$A = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_k \end{pmatrix}.$$

Here  $A_i$  are square matrices of orders  $n_i$ ,  $i = 1, \dots, k$ . Then  $A$  is a square matrix of order  $n = n_1 + \cdots + n_k$ . It is called a *block-diagonal matrix*.

It is sometimes convenient to notate matrix multiplication in block form. We shall consider only the case of two square matrices of order  $n$ , broken into blocks of the form (2.54) all of the same size:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}. \quad (2.55)$$

Here  $A_{11}$  and  $B_{11}$  are square matrices of order  $p$ ,  $A_{12}$  and  $B_{12}$  are matrices of type  $(p, n - p)$ ,  $A_{21}$  and  $B_{21}$  are matrices of type  $(n - p, p)$ ,  $A_{22}$  and  $B_{22}$  are square matrices of order  $n - p$ . Then the product  $C = AB$  is well defined and is a matrix of order  $n$  that can be broken into the same type of blocks:

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}.$$

We claim that in this case,

$$\begin{aligned} C_{11} &= A_{11}B_{11} + A_{12}B_{21}, & C_{12} &= A_{11}B_{12} + A_{12}B_{22}, \\ C_{21} &= A_{21}B_{11} + A_{22}B_{21}, & C_{22} &= A_{21}B_{12} + A_{22}B_{22}. \end{aligned} \quad (2.56)$$

In other words, the matrices (2.55) are multiplied just like matrices of order 2, except that their elements are not numbers, but blocks, that is, they are themselves matrices. The proof of formulas (2.56) follows at once from formulas (2.46). For example, let  $C = (c_{ij})$ , where  $1 \leq i, j \leq p$ . In formula (2.46), the sum of the first  $p$  terms gives the element  $c'_{ij}$  in the matrix  $A_{11}B_{11}$ , while the sum of the remaining  $n - p$  terms gives the elements  $c''_{ij}$  in the matrix  $A_{12}B_{21}$ . Of course, analogous formulas hold as well (with the same proof) for the multiplication of rectangular matrices with differing decompositions into blocks; it is necessary only that these partitions agree among themselves in such a way that the products of all matrices appearing in the formulas are defined. However, in what follows, only the case (2.55) described above will be necessary.

The transpose operation is connected with multiplication by an important relationship. Let the matrix  $A$  be of type  $(n, m)$ , and matrix  $B$  of type  $(m, k)$ . Then

$$(AB)^* = B^*A^*. \quad (2.57)$$

Indeed, by the definition of matrix product (formula (2.46)), an element of the matrix  $AB$  standing at the intersection of the  $j$ th row and  $i$ th column is equal to

$$a_{j1}b_{1i} + a_{j2}b_{2i} + \cdots + a_{jm}b_{mi}, \quad \text{where } i = 1, \dots, n, j = 1, \dots, k. \quad (2.58)$$

By definition of the transpose, the expression (2.58) gives us the value of the element of the matrix  $(AB)^*$  standing at the intersection of the  $i$ th row and the  $j$ th column. On the other hand, let us consider the product of matrices  $B^*$  and  $A^*$ , using the rule “row times column” formulated above. Then, taking into account the definition of the transpose, we obtain that the element of the matrix  $B^*A^*$  standing at the intersection of the  $i$ th row and  $j$ th column is equal to the product of the  $i$ th column of the matrix  $B$  and the  $j$ th row of the matrix  $A$ , that is, equal to

$$b_{1i}a_{j1} + b_{2i}a_{j2} + \cdots + b_{mi}a_{jm}.$$

This expression coincides with the formula (2.58) for the element of the matrix  $(AB)^*$  standing at the corresponding place, and this establishes equality (2.57).

It is possible to express, using the operation of multiplication, the elementary transformations of matrices that we used in Sect. 1.2 in studying systems of linear equations. Without specifying this especially, we shall continue to keep in mind that we are always multiplying matrices whose product is well defined.

Suppose that we are given a rectangular matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

Let us consider a square matrix of order  $m$  obtained from the identity matrix of order  $m$  by interchanging the  $i$ th and  $j$ th rows:

$$T_{ij} = \begin{pmatrix} 1 & & & 0 & & & & & & \\ & \ddots & & \vdots & & & & & & \\ & & 1 & \vdots & & & \boxed{j} & & & \\ & & & \vdots & & & \downarrow & & & \\ 0 & \cdots & \cdots & \mathbf{0} & 0 & \cdots & 0 & \mathbf{1} & \cdots & \cdots & 0 & \leftarrow \boxed{i} \\ & & & 0 & 1 & & & 0 & & & \\ & & & \vdots & & \ddots & & \vdots & & & \\ & & & 0 & & & 1 & 0 & & & \\ 0 & \cdots & \cdots & \mathbf{1} & 0 & \cdots & 0 & \mathbf{0} & \cdots & \cdots & 0 & \leftarrow \boxed{j} \\ & & & \uparrow & & & & \vdots & 1 & & \\ & & & \boxed{i} & & & & \vdots & & & \\ & & & & & & & \vdots & & \ddots & \\ & & & & & & & 0 & & & 1 \end{pmatrix}.$$

An easy check shows that  $T_{ij}A$  is also obtained from  $A$  by transposing the  $i$ th and  $j$ th rows. Therefore, we can express an elementary operation of type I on a matrix  $A$  by multiplication on the left by a suitable matrix  $T_{ij}$ .

Let us consider (for  $i \neq j$ ) a square matrix  $U_{ij}(c)$  of order  $m$  depending on the number  $c$ :

$$U_{ij}(c) = \begin{pmatrix} 1 & & & 0 & & & & & & \\ & \ddots & & \vdots & & & & & & \\ & & 1 & \vdots & & & \boxed{j} & & & \\ & & & \vdots & & & \downarrow & & & \\ 0 & \cdots & \cdots & \mathbf{1} & 0 & \cdots & 0 & \mathbf{c} & \cdots & \cdots & 0 & \leftarrow \boxed{i} \\ & & & 0 & 1 & & & 0 & & & \\ & & & \vdots & & \ddots & & \vdots & & & \\ & & & 0 & & & 1 & 0 & & & \\ 0 & \cdots & \cdots & \mathbf{0} & 0 & \cdots & 0 & \mathbf{1} & \cdots & \cdots & 0 & \leftarrow \boxed{j} \\ & & & \uparrow & & & & \vdots & 1 & & \\ & & & \boxed{i} & & & & \vdots & & & \\ & & & & & & & \vdots & & \ddots & \\ & & & & & & & 0 & & & 1 \end{pmatrix}. \quad (2.59)$$

It is obtained from the identity matrix of order  $m$  by adding the  $j$ th row multiplied by  $c$  to the  $i$ th row. An equally easy verification shows that the matrix  $U_{ij}(c)A$  is obtained from  $A$  by adding the  $j$ th row multiplied by the number  $c$  to the  $i$ th row. Therefore, we can also write an elementary operation of type II in terms of matrix multiplication. Consequently, Theorem 1.15 in matrix form can be expressed as follows:

**Theorem 2.53** *An arbitrary matrix  $A$  of type  $(m, n)$  can be brought into echelon form by multiplying on the left by the product of a number of suitable matrices  $T_{ij}$  and  $U_{ij}(c)$  (in the proper order).*

Let us examine the important case in which  $A$  and  $B$  are square matrices of order  $n$ . Then their product  $C = AB$  is also a square matrix of order  $n$ .

**Theorem 2.54** *The determinant of the product of two square matrices of identical orders is equal to the product of their determinants. That is,  $|AB| = |A| \cdot |B|$ .*

*Proof* Let us consider the determinant  $|AB|$  for a fixed matrix  $B$  as a function, which we denote by  $F(A)$ , of the rows of the matrix  $A$ . We shall prove first that the function  $F(A)$  is multilinear. We know (by Property 2.4 from Sect. 2.2) that the determinant  $|C| = F(A)$ , considered as a function of the rows of the matrix  $C = AB$ , is multilinear. In particular, it is a linear function of the  $i$ th row of the matrix  $C$ , that is,

$$F(A) = \alpha_1 c_{i1} + \alpha_2 c_{i2} + \cdots + \alpha_n c_{in} \quad (2.60)$$

for some numbers  $\alpha_1, \dots, \alpha_n$ . Let us focus attention on the fact that according to formula (2.46), the  $i$ th row of the matrix  $C = AB$  depends only on the  $i$ th row of the matrix  $A$ , while the remaining rows of the matrix  $C$ , in contrast, do not depend on this row. After substituting into formula (2.60) the expressions (2.46) for the elements of the  $i$ th row and collecting like terms, we obtain an expression for  $F(A)$  as a linear function of the  $i$ th row of the matrix  $A$ . Therefore, the function  $F(A)$  is multilinear in the rows of  $A$ . Now let us transpose two rows of the matrix  $A$ , say with indices  $i_1$  and  $i_2$ . Formula (2.46) shows us that the  $l$ th row of the matrix  $C$  for  $l \neq i_1, i_2$  does not change, but its  $i_1$ th and  $i_2$ th rows exchange places. Therefore,  $|C|$  changes sign. This means that the function  $F(A)$  is antisymmetric with respect to the rows of the matrix  $A$ . We can apply to this function Theorem 2.15, and we then obtain that  $F(A) = k|A|$ , where  $k = F(E) = |EB| = |B|$ , since for an arbitrary matrix  $B$ , the relationship  $EB = B$  is satisfied. We thereby obtain the equality  $F(A) = |A| \cdot |B|$ , whence according to our definition,  $F(A) = |AB|$ .  $\square$

Theorem 2.54 has a beautiful generalization to rectangular matrices known as the *Cauchy–Binet identity*. We shall not prove it at present, but shall give only its formulation (a natural proof will be given in Sect. 10.5 on p. 377).

The product of two rectangular matrices  $B$  and  $A$  results in a square matrix of order  $m$  if  $B$  is of type  $(m, n)$ , and  $A$  is of type  $(n, m)$ . The minors of the matrices  $B$

and  $A$  of the same order equal to the lesser of  $n$  and  $m$  are called *associates* if they stand in the columns (of matrix  $B$ ) and rows (of matrix  $A$ ) with the same indices. The Cauchy–Binet identity asserts that the determinant  $|BA|$  is equal to 0 if  $n < m$ , and  $|BA|$  is equal to the sum of the associated minors of order  $m$  if  $n \geq m$ . In this case, the sum is taken over all collections of rows (of matrix  $A$ ) and columns (of matrix  $B$ ) with increasing indices  $i_1 < i_2 < \cdots < i_m$ .

We have a beautiful special case of the Cauchy–Binet identity when

$$B = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{pmatrix}, \quad A = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \vdots \\ a_n & b_n \end{pmatrix}.$$

Then

$$BA = \begin{pmatrix} a_1^2 + a_2^2 + \cdots + a_n^2 & a_1b_1 + a_2b_2 + \cdots + a_nb_n \\ a_1b_1 + a_2b_2 + \cdots + a_nb_n & b_1^2 + b_2^2 + \cdots + b_n^2 \end{pmatrix},$$

and the associated minors assume the form

$$\begin{vmatrix} a_i & b_i \\ a_j & b_j \end{vmatrix}$$

for all  $i < j$ , taking values from 1 to  $n$ . The Cauchy–Binet identity gives us the equality

$$\begin{aligned} & (a_1^2 + a_2^2 + \cdots + a_n^2)(b_1^2 + b_2^2 + \cdots + b_n^2) - (a_1b_1 + a_2b_2 + \cdots + a_nb_n)^2 \\ &= \sum_{i < j} (a_ib_j - a_jb_i)^2. \end{aligned}$$

In particular, we derive from it the well-known inequality

$$(a_1^2 + a_2^2 + \cdots + a_n^2)(b_1^2 + b_2^2 + \cdots + b_n^2) \geq (a_1b_1 + a_2b_2 + \cdots + a_nb_n)^2.$$

The operations of addition and multiplication of matrices make it possible to define *polynomials* in matrices. In this we shall of course assume that we are always speaking about square matrices of a certain fixed order. We shall first define the operation of *exponentiation*, namely raising a matrix to the  $n$ th power. By definition,  $A^n$  for  $n > 0$  is the result of multiplying the matrix  $A$  by itself  $n$  times, while for  $n = 0$ , the result will be the identity matrix  $E$ .

**Definition 2.55** Let  $f(x) = \alpha_0 + \alpha_1x + \cdots + \alpha_kx^k$  be a polynomial with numeric coefficients. Then a *matrix polynomial*  $f$  for a matrix  $A$  is the matrix

$$f(A) = \alpha_0E + \alpha_1A + \cdots + \alpha_kA^k.$$

Let us establish some simple properties of matrix polynomials.



**Lemma 2.56** *If  $f(x) + g(x) = u(x)$  and  $f(x)g(x) = v(x)$ , then for an arbitrary square matrix  $A$  we have*

$$f(A) + g(A) = u(A), \quad (2.61)$$

$$f(A)g(A) = v(A). \quad (2.62)$$

*Proof* Let  $f(x) = \sum_{i=0}^n \alpha_i x^i$  and  $g(x) = \sum_{j=0}^m \beta_j x^j$ . Then  $u(x) = \sum_r \gamma_r x^r$  and  $v(x) = \sum_s \delta_s x^s$ , where the coefficients  $\gamma_r$  and  $\delta_s$  can be written in the form

$$\gamma_r = \alpha_r + \beta_r, \quad \delta_s = \sum_{i=0}^s \alpha_i \beta_{s-i},$$

where  $\alpha_r = 0$  if  $r > n$ , and  $\beta_r = 0$  if  $r > m$ . The equality (2.61) is now perfectly obvious. For the proof of (2.62), we observe that

$$f(A)g(A) = \sum_{i=1}^n \alpha_i A^i \cdot \sum_{j=1}^n \beta_j A^j = \sum_{i,j} \alpha_i \beta_j A^{i+j}.$$

Collecting all terms for which  $i + j = s$ , we obtain formula (2.62).  $\square$

**Corollary 2.57** *The polynomials  $f(A)$  and  $g(A)$  for the same matrix  $A$  commute:  $f(A)g(A) = g(A)f(A)$ .*

*Proof* The result follows from formula (2.62) and the equality  $f(x)g(x) = g(x)f(x)$ .  $\square$

Let us observe that the analogous assertion to the lemma just proved is not true for polynomials in several variables. For example, the identity  $(x + y)(x - y) = x^2 - y^2$  will not be preserved in general if we replace  $x$  and  $y$  with arbitrary matrices. The reason for this is that the identity depends on the relationship  $xy = yx$ , which does not hold for arbitrary matrices.

## 2.10 Inverse Matrices

In this section we shall consider exclusively square matrices of a given order  $n$ .

**Definition 2.58** A matrix  $B$  is called the *inverse* of the matrix  $A$  if

$$AB = E. \quad (2.63)$$

Here  $E$  denotes the identity matrix of the fixed order  $n$ .

Not every matrix has an inverse. Indeed, applying Theorem 2.54 on the determinant of a matrix product to equality (2.63), we obtain

$$|E| = |AB| = |A| \cdot |B|,$$

and since  $|E| = 1$ , then we must have  $|A| \cdot |B| = 1$ . Clearly, such a relationship is impossible if  $|A| = 0$ . Therefore, no singular matrix can have an inverse. The following theorem shows that the converse of this statement is also true.

**Theorem 2.59** *For every nonsingular matrix  $A$  there exists a matrix  $B$  satisfying the relationship (2.63).*

*Proof* Let us denote the yet unknown  $j$ th column of the desired inverse matrix  $B$  by  $[b]_j$ , while  $[e]_j$  will denote the  $j$ th column of the identity matrix  $E$ . The columns  $[b]_j$  and  $[e]_j$  are matrices of type  $(n, 1)$ , and by the product rule for matrices, the equality (2.63) is equivalent to the  $n$  relationships

$$A[b]_j = [e]_j, \quad j = 1, \dots, n. \quad (2.64)$$

Therefore, it suffices to prove the solvability of each (for each fixed  $j$ ) system of linear equations (2.64) for the  $n$  unknowns that are the elements of the matrix  $B$  appearing in column  $[b]_j$ . But for every index  $j$ , the matrix of this system is  $A$ , and by hypothesis,  $|A| \neq 0$ . By Theorem 2.12, such a system has a solution (and indeed, a unique one). Taking the solution of the system obtained for each index  $j$  as the  $j$ th column of the matrix  $B$ , we obtain a matrix satisfying the condition (2.63), that is, we have found an inverse to the matrix  $A$ .  $\square$

Let us recall that matrix multiplication is not commutative, that is, in general,  $AB \neq BA$ . Therefore, it would be natural to consider another possible definition of the inverse matrix of  $A$ , namely a matrix  $C$  such that

$$CA = E. \quad (2.65)$$

The same reasoning as that carried out at the beginning of this section shows that such a matrix  $C$  does not exist if  $A$  is singular.

**Theorem 2.60** *For an arbitrary nonsingular matrix  $A$ , there exists a matrix  $C$  satisfying relationship (2.65).*

*Proof* This theorem can be proved in two different ways. First, it would be possible to repeat in full the proof of Theorem 2.59, considering now instead of the columns of the matrices  $C$  and  $E$ , their rows. But perhaps there is a somewhat more elegant proof that derives Theorem 2.60 directly from Theorem 2.59. To this end, let us apply Theorem 2.59 to the transpose matrix  $A^*$ . By Theorem 2.32,  $|A^*| = |A|$ , and therefore,  $|A^*| \neq 0$ , which means that there exists a matrix  $B$  such that

$$A^*B = E. \quad (2.66)$$

Let us apply the transpose operation to both sides of (2.66). It is clear that  $E^* = E$ . On the other hand, by (2.57),

$$(A^*B)^* = B^*(A^*)^*,$$

and it is easily verified that  $(A^*)^* = A$ . We therefore obtain  $B^*A = E$ , and in (2.65) we can take the matrix  $B^*$  for  $C$ , where  $B$  is defined by (2.66).  $\square$

The matrices  $B$  from (2.63) and  $C$  from (2.65) can make equal claim to the title of inverse of the matrix  $A$ . Fortunately, we do not obtain here two different definitions of the inverse, since these two matrices coincide. Namely, we have the following result.

**Theorem 2.61** *For any nonsingular matrix  $A$  there exists a unique matrix  $B$  satisfying (2.63) and a unique matrix  $C$  satisfying (2.65). Moreover, the two matrices are equal.*

*Proof* Let  $A$  be a nonsingular matrix. We shall show that the matrix  $B$  satisfying (2.63) is unique. Let us assume that there exists another matrix,  $B'$ , such that  $AB' = E$ . Then  $AB = AB'$ , and if we multiply both sides of this equality by the matrix  $C$  such that  $CA = E$ , whose existence is guaranteed by Theorem 2.60, then by the associative property of matrix multiplication, we obtain  $(CA)B = (CA)B'$ , whence follows the equality  $EB = EB'$ , that is,  $B = B'$ . In exactly the same way we can prove the uniqueness of  $C$  satisfying (2.65).

Now let us show that  $B = C$ . To this end, we consider the product  $C(AB)$  and make use of the associative property of multiplication:

$$C(AB) = (CA)B. \quad (2.67)$$

Then on the one hand,  $AB = E$  and  $C(AB) = CE = C$ , while on the other hand,  $CA = E$  and  $(CA)B = EB = B$ , and relationship (2.67) gives us  $B = C$ .  $\square$

This unique (by Theorem 2.61) matrix  $B = C$  is denoted by  $A^{-1}$  and is called *the* inverse of the matrix  $A$ . Thus for every nonsingular matrix  $A$ , there exists an inverse matrix  $A^{-1}$  satisfying the relationship

$$AA^{-1} = A^{-1}A = E, \quad (2.68)$$

and such a matrix  $A^{-1}$  is unique.

In following the proof of Theorem 2.59, we see that it is possible to derive an explicit formula for the inverse matrix. We again assume that the matrix  $A$  is nonsingular, and following the notation used in the proof of Theorem 2.59, we arrive at the system of equations (2.64). Since  $|A| \neq 0$ , we can find a solution of this system using Cramer's rule (2.35). For an arbitrary index  $j = 1, \dots, n$  in system (2.64), the

$i$ th unknown coincides with the element  $b_{ij}$  of the matrix  $B$ . Using Cramer's rule, we obtain for it the value

$$b_{ij} = \frac{D_{ij}}{|A|}, \quad (2.69)$$

where  $D_{ij}$  is the determinant of the matrix obtained from  $A$  by replacing the  $i$ th column by the column  $[e]_j$ . The determinant  $D_{ij}$  can be expanded along the  $i$ th column, and by formula (2.30), we obtain that it is equal to the cofactor of the unique nonzero (and equal to 1) element of the  $i$ th column. Since the  $i$ th column is equal to  $[e]_j$ , there is a 1 at the intersection of the  $i$ th column (which we replaced by  $[e]_j$ ) and the  $j$ th row. Therefore,  $D_{ij} = A_{ji}$ , and formula (2.69) yields

$$b_{ij} = \frac{A_{ji}}{|A|}.$$

This is an explicit formula for the elements of the inverse matrix. In words, this can be formulated thus: to obtain the inverse matrix of a nonsingular matrix  $A$ , one must replace every element with its cofactor, then transpose the matrix thus obtained and multiply it by the number  $|A|^{-1}$ .

For example, for the  $2 \times 2$  matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with  $\delta = |A| = ad - bc \neq 0$ , we obtain the inverse matrix

$$A^{-1} = \begin{pmatrix} d/\delta & -b/\delta \\ -c/\delta & a/\delta \end{pmatrix}.$$

The concept of inverse matrix provides a simple and elegant notation for the solution of a system of  $n$  equations in  $n$  unknowns. If, as in the previous section, we write down the system of linear equations (1.3) with  $n = m$  and  $A$  a nonsingular matrix in the form  $A[\mathbf{x}] = [\mathbf{b}]$ , where  $[\mathbf{x}]$  is the column of unknowns  $x_1, \dots, x_n$  and  $[\mathbf{b}]$  is the column consisting of the constants of the system, then multiplying this relationship on the left by the matrix  $A^{-1}$ , we obtain the solution in the form  $[\mathbf{x}] = A^{-1}[\mathbf{b}]$ . Thus, in matrix notation, the formulas for the solution of a system of  $n$  linear equations in  $n$  unknowns look just like those for a single equation in a single unknown. But if we use the formulas for the inverse matrix, then we see that the relationship  $[\mathbf{x}] = A^{-1}[\mathbf{b}]$  exactly coincides with Cramer's rule, so that this more elegant notation gives us nothing essentially new.

Let us consider the matrix  $\bar{A} = (\bar{a}_{ij})$ , in which the element  $\bar{a}_{ij} = A_{ji}$  is the cofactor of the element  $a_{ji}$  of the matrix  $A$ . The matrix  $\bar{A}$  is called the *adjugate* matrix to  $A$ . For a matrix  $A$  of order  $n$ , the elements of the adjugate matrix are polynomials of degree  $n - 1$  in the elements of  $A$ . Formula (2.69) for the inverse matrix shows that

$$A\bar{A} = \bar{A}A = |A|E. \quad (2.70)$$

The advantage of the adjugate matrix  $\overline{A}$  compared to the inverse matrix  $A^{-1}$  is that the definition of  $\overline{A}$  does not require division by  $|A|$ , and formula (2.70), in contrast to the analogous formula (2.68), holds even for  $|A| = 0$ , that is, even for singular square matrices, as the proof of Cramer's rule demonstrates. We shall make use of this fact in the sequel.

In conclusion, let us return once more to the question of presenting elementary operations in terms of matrix multiplication, which we began to examine in the previous section. It is easy to see that the matrices  $T_{ij}$  and  $U_{ij}(c)$  introduced there are nonsingular, and moreover,

$$T_{ij}^{-1} = T_{ji}, \quad U_{ij}^{-1}(c) = U_{ij}(-c).$$

Therefore, Theorem 2.53 can be reformulated as follows: An arbitrary matrix  $A$  can be obtained from a particular echelon matrix  $A'$  by multiplying it on the left by matrices  $T_{ij}$  and  $U_{ij}(c)$  in a certain order.

Let us apply this result to nonsingular square matrices of order  $n$ . Since  $|T_{ij}| \neq 0$ ,  $|U_{ij}(c)| \neq 0$ , and  $|A| \neq 0$  (by assumption), the matrix  $A'$  must also be nonsingular. But a nonsingular square echelon matrix is in upper triangular form, that is, all of its elements below the main diagonal are equal to zero, namely,

$$A' = \begin{pmatrix} a'_{11} & a'_{12} & a'_{13} & \cdots & a'_{1n} \\ 0 & a'_{22} & a'_{23} & \cdots & a'_{2n} \\ 0 & 0 & a'_{33} & \cdots & a'_{3n} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a'_{nn} \end{pmatrix},$$

and moreover,  $|A'| = a'_{11}a'_{22} \cdots a'_{nn}$ . Therefore, all the elements  $a'_{11}, \dots, a'_{nn}$  on the main diagonal are different from zero.

But this matrix  $A'$  can be brought into a yet simpler form with the help of elementary operations of type II only. Namely, since  $a'_{nn} \neq 0$ , one can subtract from the rows with indices  $n-1, n-2, \dots, 1$  of the matrix  $A'$  the last row multiplied by factors that make all the elements of the  $n$ th column (except for  $a'_{nn}$ ) equal to zero. Since  $a'_{n-1n-1} \neq 0$ , it is possible in the same way to reduce to zero all elements of the  $(n-1)$ st column (except for the element  $a'_{n-1n-1}$ ). Doing this  $n$  times, we shall make all of the elements of the matrix equal to zero except those on the main diagonal. That is, we end up with the matrix

$$D = \begin{pmatrix} a'_{11} & 0 & 0 & \cdots & 0 \\ 0 & a'_{22} & 0 & \cdots & 0 \\ 0 & 0 & a'_{33} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a'_{nn} \end{pmatrix}. \quad (2.71)$$

A matrix all of whose elements are equal to zero except for those on the main diagonal is called a *diagonal matrix*. We have thus proved that a matrix  $A'$  can be

obtained from a diagonal matrix  $D$  by multiplying it on the left by matrices of the form  $T_{ij}$  and  $U_{ij}(c)$  in some order.

Let us note that multiplication by a matrix  $T_{ij}$  (that is, an elementary operation of type I) can be replaced by multiplication on the left by matrices of type  $U_{ij}(c)$  for various  $c$  and by a certain simpler matrix. Namely, the interchange of the  $i$ th and  $j$ th rows can be obtained using the following four operations:

1. Addition of the  $i$ th row to the  $j$ th row.
2. Subtraction of the  $j$ th row from the  $i$ th row.
3. Addition of the  $i$ th row to the  $j$ th row.

Schematically, this can be depicted as follows, where the  $i$ th and  $j$ th rows are denoted by  $\mathbf{c}_i$  and  $\mathbf{c}_j$ :

$$\begin{pmatrix} \mathbf{c}_i \\ \mathbf{c}_j \end{pmatrix} \xrightarrow{1} \begin{pmatrix} \mathbf{c}_i \\ \mathbf{c}_i + \mathbf{c}_j \end{pmatrix} \xrightarrow{2} \begin{pmatrix} -\mathbf{c}_j \\ \mathbf{c}_i + \mathbf{c}_j \end{pmatrix} \xrightarrow{3} \begin{pmatrix} -\mathbf{c}_j \\ \mathbf{c}_i \end{pmatrix}.$$

4. It is now necessary to introduce a new type of operation: its effect is to multiply the  $i$ th row by  $-1$  and is achieved by multiplying (with  $k = i$ ) our matrix on the left by the square matrix

$$S_k = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & \boxed{k} & & \\ & & \downarrow & -1 & \leftarrow \boxed{k} & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix}, \quad (2.72)$$

where there is  $-1$  at the intersection of the  $k$ th row and  $k$ th column.

We may now reformulate Theorem 2.53 as follows:

**Theorem 2.62** *Any nonsingular matrix can be obtained from a diagonal matrix by multiplying it on the left by certain matrices  $U_{ij}(c)$  of the form (2.59) and matrices  $S_k$  of the form (2.72).*

We shall use this result in Sect. 4.4 when we introduce the *orientation* of a real vector space. Furthermore, Theorem 2.62 provides a simple and convenient method of computing the inverse matrix, in a manner based on Gaussian elimination. To this end, we introduce yet another (a third) type of elementary matrix operation, which consists in multiplying the  $k$ th row of a matrix by an arbitrary nonzero number  $\alpha$ . It is clear that the result of such an operation can be obtained by multiplying our

matrix on the left by the square matrix

$$V_k(\alpha) = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & \boxed{k} & \\ & & & \downarrow & \\ & & & \alpha & \leftarrow \boxed{k} \\ & & & & 1 \\ & & & & & \ddots \\ & & & & & & 1 \end{pmatrix}, \quad (2.73)$$

where the number  $\alpha$  stands at the intersection of the  $k$ th row and  $k$ th column. By multiplying the matrix (2.71) on the left by the matrices  $V_1(a'_{11}^{-1}), \dots, V_n(a'_{nn}^{-1})$ , we transform it into the identity matrix.

From Theorem 2.62, it follows that every nonsingular matrix can be obtained from the identity matrix by multiplying it on the left by matrices  $U_{ij}(c)$  of the type given in (2.59), matrices  $S_k$  from (2.72), and matrices  $V_k(\alpha)$  of the form of (2.73). However, since multiplication by each of these matrices is equivalent to an elementary operation of one of the three types, this means that every nonsingular matrix can be obtained from the identity matrix using a sequence of such operations, and conversely, using a certain number of elementary operations of all three types, it is possible to obtain the identity from an arbitrary nonsingular matrix. This gives us a convenient method of computing the inverse matrix. Indeed, suppose that using some sequence of elementary operations of all three types, we have transformed matrix  $A$  to the identity matrix  $E$ . Let us denote by  $B$  the product of all the matrices  $U_{ij}(c)$ ,  $S_k$ , and  $V_k(\alpha)$ , whose product corresponds to the given operations (in the obvious order: the matrix representing each successive operation stands to the left of the previous one). Then  $BA = E$ , from which it follows that  $B = A^{-1}$ . Then after applying the same sequence of elementary operations to the matrix  $E$ , we obtain from it the matrix  $BE = B$ , that is,  $A^{-1}$ . Therefore, to compute  $A^{-1}$ , it suffices to transform the matrix  $A$  to  $E$  using elementary operations of the three types (as was shown above), while simultaneously applying the same operations to the matrix  $E$ . The matrix obtained from  $E$  as a result of the same elementary operations will be  $A^{-1}$ .

Let  $C$  be an arbitrary matrix of type  $(m, n)$ . We shall show that for an arbitrary nonsingular square matrix  $A$  of order  $m$ , the rank of the product  $AC$  is equal to the rank of  $C$ . Indeed, as we have already seen, the matrix  $A$  can be transformed into  $E$  by applying some sequence of elementary operations of the three types to its rows, to which corresponds multiplication on the left by the matrix  $A^{-1}$ . Applying the same sequence of operations to  $AC$ , we clearly obtain the matrix  $A^{-1}AC = C$ . By Theorem 2.37, the rank of a matrix is not changed by elementary operations of types I and II. It also does not change under elementary operations of type III. This clearly follows from the fact that every minor is a linear function of its rows, and consequently, every nonzero minor of a matrix remains a nonzero minor after multiplication of any of its rows by an arbitrary nonzero number. Therefore, the rank of the matrix  $AC$  is equal to the rank of  $C$ .

Using an analogous argument for the columns as was given for the rows, or simply using Theorem 2.36, we obtain the following useful result.

**Theorem 2.63** *For any matrix  $C$  of type  $(m, n)$  and any nonsingular square matrices  $A$  and  $B$  of orders  $m$  and  $n$ , the rank of  $ACB$  is equal to the rank of  $C$ .*



## Chapter 3

# Vector Spaces

### 3.1 The Definition of a Vector Space

Vectors on a line, in the plane, or in space play a significant role in mathematics, and especially in physics. Vectors represent the displacement of bodies, or their speed, acceleration, or the force applied to them, among many other things.

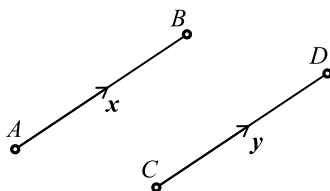
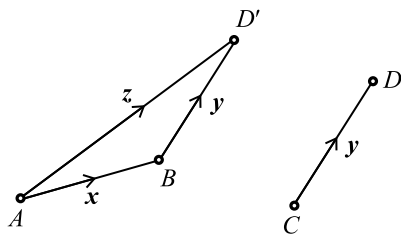
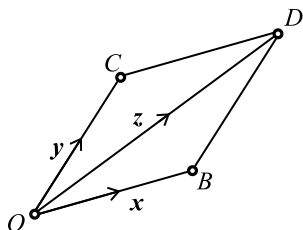
In a course in elementary mathematics or physics, a *vector* is defined as a directed line segment. The word *directed* indicates that a direction is assigned to the segment, often indicated by an arrow drawn above it. Or else, perhaps, one of the two endpoints of the segment  $[A, B]$ , say  $A$ , is called the *beginning*, while the other,  $B$ , is the *end*, and then the direction is given as motion from the beginning of the segment to the end. Then two vectors  $\mathbf{x} = \overrightarrow{AB}$  and  $\mathbf{y} = \overrightarrow{CD}$  are said to be *equal* if it is possible by means of parallel translation to join the segments  $\mathbf{x}$  and  $\mathbf{y}$  in such a way that the beginning  $A$  of segment  $\mathbf{x}$  coincides with the beginning  $C$  of segment  $\mathbf{y}$  (in which case their ends must coincide as well); see Fig. 3.1.

The fact that we consider the two different vectors in the figure to be equal does not represent anything unusual in mathematics or generally in human thought. Rather, it represents the usual method of *abstraction*, whereby we focus our attention on some important property of the objects under consideration. Thus in geometry, we consider certain triangles to be equal, even though they are drawn on different sheets of paper. Or in arithmetic, we might consider equal the number of people in a boat and the number of apples on a tree.

It is obvious that having chosen a certain point  $O$  (on a line, in the plane, or in space), we can find a vector (indeed the unique one) equal to a given vector  $\mathbf{x}$  whose beginning coincides with the point  $O$ .

The laws of addition of velocities, accelerations, and forces lead to the following definition of vector addition. The *sum* of vectors  $\mathbf{x} = \overrightarrow{AB}$  and  $\mathbf{y} = \overrightarrow{CD}$  is the vector  $\mathbf{z} = \overrightarrow{AD'}$ , where  $D'$  is the end of vector  $\overrightarrow{BD'}$ , a vector equal to  $\mathbf{y}$  whose beginning coincides with the end  $B$  of the vector  $\mathbf{x}$ ; see Fig. 3.2.

If we replace all of these vectors with equal vectors but having as their beginning the fixed point  $O$ , then vector addition will proceed by the well-known “parallelogram law”; see Fig. 3.3.

**Fig. 3.1** Equal vectors**Fig. 3.2** Vector summation**Fig. 3.3** The parallelogram law

There is also a definition of multiplication of a vector  $x$  by a number  $\alpha$ . For now, in speaking about numbers, we shall mean real numbers (we shall have something to say later about the more general situation). If  $\alpha > 0$  and  $x$  is the vector  $\overrightarrow{AB}$ , then the *product*  $\alpha x$  is defined to be the vector  $\overrightarrow{AC}$  lying on the same line as  $[A, B]$  in such a way that the point  $C$  lies on the same side of  $A$  as the point  $B$  and such that the segment  $[A, C]$  is  $\alpha$  times the length of the segment  $[A, B]$ . (Note that if  $\alpha < 1$ , then the segment  $[A, C]$  is shorter than the segment  $[A, B]$ .) Denoting by  $|AB|$  the length of the segment  $[A, B]$ , we shall express this by way of the formula  $|AC| = \alpha|AB|$ . However, if  $\alpha < 0$  and  $\alpha = -\beta$ , where then  $\beta > 0$ , then the *product*  $\alpha x$  is defined to be the vector  $\overrightarrow{CA}$ , where  $\beta x = \overrightarrow{AC}$ .

We shall not derive the simple properties of vector addition and multiplication of a vector by a number. We observe only that they are amazingly similar for vectors on a line, in the plane, and in space. This similarity indicates that we are dealing only with a special case of a general concept. In this and several subsequent chapters, we shall present the theory of vectors and the spaces consisting of them of arbitrary dimension  $n$  (including even some facts relating to spaces whose dimension is infinite).

How do we formulate such a definition? In the case of vectors on a line, in the plane, and in space, we shall use the intuitively clear concept of directed line seg-

ment. But what if we are not convinced that our interlocutor shares the same intuition? For example, suppose we wanted to share our knowledge with an extraterrestrial with whom we are communicating by radio?

A technique was long ago devised for overcoming such difficulties in the sciences. It involves defining (or in our terminology, reporting to the extraterrestrial) not *what are* the objects under consideration (vectors, etc.), but the *relationships between them*, or in other words, their *properties*. For example, in geometry, one leaves undefined such notions as point, line, and the property of a line passing through a point, and instead formulates some of their properties, for instance that between two distinct points there passes one and only one line. Such a method of defining new concepts is called *axiomatic*. In this course on linear algebra, the vector space will be the first object to be defined axiomatically. Till now, new concepts have been defined using constructions or formulas, such as the definition of the determinant of a matrix (defined either inductively, using the rule of expansion by columns, or derived using the rather complicated explicit formula (2.44) from Sect. 2.7). It is, however, possible that the reader has encountered the concepts of groups and fields, which are also defined axiomatically, but may not have investigated them in detail, in contrast to the notion of a vector space, the study of which will occupy this entire chapter.

With that, we move on to the definition of a vector space.

**Definition 3.1** A *vector* (or *linear*) *space* is a set  $L$  (whose elements we shall call *vectors* and denote by  $\mathbf{x}$ ,  $\mathbf{y}$ ,  $\mathbf{z}$ , etc.) for which the following conditions are satisfied:

- (1) There is a rule for associating with any two vectors  $\mathbf{x}$  and  $\mathbf{y}$  a third vector, called their *sum* and denoted by  $\mathbf{x} + \mathbf{y}$ .
- (2) There is a rule for associating with any vector  $\mathbf{x}$  and any number  $\alpha$  a new vector, called the *product* of  $\alpha$  and  $\mathbf{x}$  and denoted by  $\alpha\mathbf{x}$ . (The numbers  $\alpha$  by which a vector can be multiplied, be they real, complex, or from any field  $\mathbb{K}$ , are called *scalars*.)

These operations must satisfy the following conditions:

- (a)  $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ .
- (b)  $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ .
- (c) There exists a vector  $\mathbf{0} \in L$  such that for an arbitrary vector  $\mathbf{x} \in L$ , the sum  $\mathbf{x} + \mathbf{0}$  is equal to  $\mathbf{x}$  (the vector  $\mathbf{0}$  is called the *null vector*).
- (d) For each vector  $\mathbf{x} \in L$ , there exists a vector  $-\mathbf{x} \in L$  such that  $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$  (the vectors  $\mathbf{x}$  and  $-\mathbf{x}$  are called *additive inverses* or *opposites* of each other).<sup>1</sup>
- (e) For an arbitrary scalar  $\alpha$  and vectors  $\mathbf{x}$  and  $\mathbf{y}$ ,

$$\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}.$$

---

<sup>1</sup>Readers who are familiar with the concept of a *group* will be able to reformulate conditions (a)–(c) in a compact way by saying that with respect to the operation of vector addition, the vectors form an *abelian group*.

(f) For arbitrary scalars  $\alpha$  and  $\beta$  and vector  $\mathbf{x}$ ,

$$(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}.$$

(g) Similarly,

$$\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}.$$

(h) For an arbitrary vector  $\mathbf{x}$ ,

$$1\mathbf{x} = \mathbf{x} \quad \text{and} \quad 0\mathbf{x} = \mathbf{0}.$$

In the last equality, the  $\mathbf{0}$  on the right-hand side denotes the null vector of the space  $L$ , while the  $0$  on the left is the scalar zero (these will always be so denoted using lighter and heavier type).

It is easy to prove that there is a unique null vector in  $L$ . Indeed, if there were another null vector  $\mathbf{0}'$ , then by definition, we would have the equality  $\mathbf{0}' = \mathbf{0}' + \mathbf{0} = \mathbf{0}$ , from which it follows that  $\mathbf{0}' = \mathbf{0}$ .

Using properties (a) through (d) and the uniqueness of the null vector, it is easily proved that for an arbitrary  $\mathbf{x}$ , there is a unique additive inverse vector  $-\mathbf{x}$  in  $L$ .

It follows from properties (g) and (h) that the vector  $-\mathbf{x}$  is obtained by multiplying the vector  $\mathbf{x}$  by the scalar  $-1$ . Indeed, since

$$\mathbf{x} + (-1)\mathbf{x} = 1\mathbf{x} + (-1)\mathbf{x} = (1 + (-1))\mathbf{x} = 0\mathbf{x} = \mathbf{0},$$

we obtain by the uniqueness of the additive inverse that  $(-1)\mathbf{x} = -\mathbf{x}$ . Analogously, from properties (f) and (h), it follows that for every vector  $\mathbf{x}$  and natural number  $k$ , the vector  $k\mathbf{x}$  is equal to the  $k$ -fold sum  $\mathbf{x} + \cdots + \mathbf{x}$ .

*Remark 3.2 (On scalars and fields)* We would like to make more precise what we mean by *scalars*  $\alpha, \beta$ , etc. in the definition of vector space above. The majority of readers will probably assume that we are talking about real numbers. In this case,  $L$  is called a *real* vector space. But those who are familiar with complex numbers may choose to understand the scalars  $\alpha, \beta$ , etc., as complex. In that case,  $L$  will be called a *complex* vector space. The theory developed below will be applicable in this case as well. Finally, the reader familiar with the concept of *field* may combine these two cases, understanding the scalars involved in the definition of a vector space to be elements of any field  $\mathbb{K}$ . Then  $L$  will be called a vector space *over the field*  $\mathbb{K}$ .

Strictly speaking, this question of scalars could have been addressed in the preceding chapters in which we discussed numbers without going into much detail. The answer would have been the same: by scalars, one may understand real numbers, complex numbers, or the elements of any field. All of our arguments apply equally to all three cases. The only exception is the proof of Property 2.10 from Sect. 2.2, in which we used the fact that from the equality  $2D = 0$  it followed that  $D = 0$ . A field

in which that assertion is true for every element  $D$  is called a field of *characteristic*<sup>2</sup> different from 2. Nonetheless, it is possible to prove that Property 2.10 holds in the general case as well.

*Example 3.3* We present here a few examples of vector spaces.

- (a) The set of vectors on a line, in the plane, or in space as we have previously discussed.
- (b) In Sect. 2.9, we introduced the notions of addition of matrices and multiplication of a matrix by a number. It is easily verified that the set of matrices of a given type  $(m, n)$  with operations thus defined is a vector space. That conditions (a) through (h) are satisfied reduces to the corresponding properties of numbers. In particular, the set of rows (or columns) of a given length  $n$  is a vector space. We shall denote this space by  $\mathbb{K}^n$  if the row (or column) elements belong to the field  $\mathbb{K}$ . Here it is understood that if we are operating with real numbers only, then  $\mathbb{K} = \mathbb{R}$ , and the field will then be denoted by  $\mathbb{R}^n$ . If we are using complex numbers, then  $\mathbb{K} = \mathbb{C}$ , and the vector space will be denoted by  $\mathbb{C}^n$ . The reader may choose any of these designations.
- (c) Let  $L$  be the set of all continuous functions defined on a given interval  $[a, b]$  taking real or complex values. We define addition of such functions and multiplication by a scalar in the usual way. It is then clear that  $L$  is a vector space.
- (d) Let  $L$  be the set of all polynomials (of arbitrary degree) with real or complex coefficients or coefficients in a field  $\mathbb{K}$ . Addition and multiplication by a scalar are defined as usual. Then it is obvious that  $L$  is a vector space.
- (e) Let  $L$  be the collection of all polynomials whose degree does not exceed a fixed number  $n$ . Everything else is the same as in the previous example. We again obtain a vector space (one for each value of  $n$ ).

**Definition 3.4** A subset  $L'$  of a vector space  $L$  is called a *subspace* of  $L$  if for arbitrary vectors  $x, y \in L'$ , their sum  $x + y$  is also in  $L'$ , and for an arbitrary scalar  $\alpha$  and vector  $x \in L'$ , the vector  $\alpha x$  is in  $L'$ .

It is obvious that  $L'$  is itself a vector space.

*Example 3.5* The space  $L$  is a subspace of itself.

*Example 3.6* The vector  $\mathbf{0}$  by itself forms a subspace. It is called the *zero space* and is denoted by  $(\mathbf{0})$ .<sup>3</sup>

---

<sup>2</sup>For readers familiar with the definition of a field, we can give a general definition: The *characteristic* of a field  $\mathbb{K}$  is the smallest natural number  $k$  such that the  $k$ -fold sum  $kD = D + \cdots + D$  is equal to 0 for every element  $D \in \mathbb{K}$  (as is easily seen, this number  $k$  is the same for all  $D \neq 0$ ). If no such natural number  $k$  exists (as in, for example, the most frequently encountered fields,  $\mathbb{K} = \mathbb{R}$  and  $\mathbb{K} = \mathbb{C}$ ), then the characteristic is defined to be zero.

<sup>3</sup>*Translator's note:* It may be tempting to consider “null space” a possible synonym for the zero space. However, that term is reserved as a synonym for “kernel,” to be introduced below, in Definition 3.67.

**Example 3.7** Consider the space encountered in analytic geometry consisting of all vectors having their beginning at a certain fixed point  $O$ . Then an arbitrary line and an arbitrary plane passing through the point  $O$  will be subspaces of the entire enclosing vector space.

**Example 3.8** Consider a system of homogeneous linear equations in  $n$  unknowns with coefficients in the field  $\mathbb{K}$ . Then the set of rows forming the solution set is a subspace  $L'$  of the space  $\mathbb{K}^n$  of rows of length  $n$ . This follows from the notation (1.10) of such a system (with  $b_i = 0$ ) and properties (1.8) and (1.9) of linear functions. The subspace  $L'$  is called the *solution subspace* of the associated system of homogeneous linear equations. The equations of the system determine the subspace  $L'$  just as the equation of a line or plane does in analytic geometry.

**Example 3.9** In the space of all polynomials, the collection of all polynomials with degree at most  $n$  (for any fixed number  $n$ ) is a subspace.

**Definition 3.10** A space  $L$  is called the *sum* of a collection of its subspaces  $L_1, L_2, \dots, L_k$  if every vector  $\mathbf{x} \in L$  can be written in the form

$$\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_k, \quad \text{where } \mathbf{x}_i \in L_i. \quad (3.1)$$

In that case, we write

$$L = L_1 + L_2 + \cdots + L_k.$$

**Definition 3.11** A space  $L$  is called the *direct sum* of its subspaces  $L_1, L_2, \dots, L_k$  if it is the sum of these subspaces and in addition, for every vector  $\mathbf{x} \in L$ , the representation (3.1) is unique. In this case, we write

$$L = L_1 \oplus L_2 \oplus \cdots \oplus L_k. \quad (3.2)$$

**Example 3.12** The space that we considered in Example 3.7 is the sum of two planes if they do not coincide; it is the sum of a line and plane if the line is not contained in the given plane; it is the sum of three lines if they do not belong to a common plane. In the second and third cases, the sum will be a direct sum. In the case of two planes, it is easily seen that the representation (3.1) is not unique. For example, we can represent the null vector as a sum of two vectors that are additive inverses of each other lying on the line that is obtained as the intersection of the two given planes.

**Example 3.13** Let us denote by  $L_i$  the vector space consisting of all monomials of degree  $i$ . Then the space  $L$  of polynomials of degree at most  $n$  can be represented as the direct sum  $L = L_0 \oplus L_1 \oplus \cdots \oplus L_n$ . This follows from the fact that an arbitrary polynomial is uniquely determined by its coefficients.

**Lemma 3.14** *Suppose the vector space  $L$  is the sum of certain of its subspaces  $L_1, L_2, \dots, L_k$ . Then in order for  $L$  to be a direct sum of these subspaces, it is necessary and sufficient that the relationship*

$$\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_k = \mathbf{0}, \quad \mathbf{x}_i \in L_i, \quad (3.3)$$

*hold only if all the  $\mathbf{x}_i$  are equal to  $\mathbf{0}$ .*

*Proof* The necessity of condition (3.3) is clear, since for the vector  $\mathbf{0} \in L$ , the equality  $\mathbf{0} = \mathbf{0} + \cdots + \mathbf{0}$ , in which the null vector of the subspace  $L_i$  stands in the  $i$ th place, is a representation of type (3.1), and the presence of another equality of the form (3.3) would contradict the definition of direct sum. To prove the sufficiency of the condition (3.3), if there are two representations (3.1),

$$\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_k, \quad \mathbf{x} = \mathbf{y}_1 + \mathbf{y}_2 + \cdots + \mathbf{y}_k,$$

then it suffices to subtract one from the other and again use the definition of direct sum.  $\square$

We observe that if  $L_1, L_2, \dots, L_k$  are subspaces of a vector space  $L$ , then their intersection  $L_1 \cap L_2 \cap \cdots \cap L_k$  is also a subspace of  $L$ , since it satisfies all the requirements in the definition of subspace. In the case  $k = 2$ , then Lemma 3.14 allows us to obtain in the following corollary another, more graphic, criterion for the sum of subspaces to be a direct sum.

**Corollary 3.15** *Suppose the vector space  $L$  is the sum of two of its subspaces  $L_1$  and  $L_2$ . Then in order that  $L$  be a direct sum, it is necessary and sufficient that one have the equality  $L_1 \cap L_2 = (\mathbf{0})$ .*

*Proof* By Lemma 3.14,  $L$  is the direct sum of its subspaces  $L_1$  and  $L_2$  if and only if the equation  $\mathbf{x}_1 + \mathbf{x}_2 = \mathbf{0}$ , where  $\mathbf{x}_1 \in L_1$  and  $\mathbf{x}_2 \in L_2$ , is satisfied only if  $\mathbf{x}_1 = \mathbf{0}$  and  $\mathbf{x}_2 = \mathbf{0}$ . But from  $\mathbf{x}_1 + \mathbf{x}_2 = \mathbf{0}$ , it follows that the vector  $\mathbf{x}_1 = -\mathbf{x}_2$  is contained in both subspaces  $L_1$  and  $L_2$ , whence it follows that it is contained in the intersection  $L_1 \cap L_2$ . Therefore, the condition  $L = L_1 \oplus L_2$  is equivalent to the satisfaction of the two conditions  $L = L_1 + L_2$  and  $L_1 \cap L_2 = (\mathbf{0})$ , which completes the proof.  $\square$

We observe that the last assertion cannot be generalized to an arbitrary number of subspaces  $L_1, \dots, L_k$ . For example, suppose that  $L$  is the plane consisting of all vectors with origin at  $O$ , and suppose that  $L_1, L_2, L_3$  are three distinct lines in this plane passing through  $O$ . It is clear that the intersection of any two of these lines consists of only the zero vector, and so a fortiori,  $L_1 \cap L_2 \cap L_3 = (\mathbf{0})$ . The plane  $L$  is the sum of its subspaces  $L_1, L_2, L_3$ , but it is not the direct sum, since it is obvious that one can produce the equality  $\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3 = \mathbf{0}$  for nonnull vectors  $\mathbf{x}_i \in L_i$ .

It is easy to see that if equality (3.2) is satisfied, then there exists a bijection between the set of vectors  $\mathbf{x} \in L$  and the set  $L_1 \times \cdots \times L_k$ , the product of the sets  $L_1, \dots, L_k$  (see the definition on page xvi). This observation provides a method for

constructing the direct sum of vector spaces that are not, so to speak, originally subspaces of a larger enclosing space and even have perhaps completely different structures from one another.

Let  $L_1, \dots, L_k$  be vector spaces. Just as for any other sets, we can define their product  $L = L_1 \times \dots \times L_k$ , which in this case is not yet a vector space. However, it is easy to make it into one by defining the sum and the product by a scalar according to the following formulas:

$$\begin{aligned}(\mathbf{x}_1, \dots, \mathbf{x}_k) + (\mathbf{y}_1, \dots, \mathbf{y}_k) &= (\mathbf{x}_1 + \mathbf{y}_1, \dots, \mathbf{x}_k + \mathbf{y}_k), \\ \alpha(\mathbf{x}_1, \dots, \mathbf{x}_k) &= (\alpha\mathbf{x}_1, \dots, \alpha\mathbf{x}_k),\end{aligned}$$

for all vectors  $\mathbf{x}_i \in L_i$ ,  $\mathbf{y}_i \in L_i$ ,  $i = 1, \dots, k$ , and an arbitrary scalar  $\alpha$ .

A simple verification shows that in this way, the definition of the operation satisfies all the conditions for the definition of a vector space, and the set  $L = L_1 \times \dots \times L_k$  becomes a vector space containing  $L_1, \dots, L_k$  among its subspaces. If we wish to be technically precise, then the subspaces of  $L$  are not the  $L_i$  themselves, but the sets  $L'_i = (\mathbf{0}) \times \dots \times L_i \times \dots \times (\mathbf{0})$ , where  $L_i$  stands in the  $i$ th place, with the zero space at all the remaining places other than  $L_i$ . However, we shall close our eyes to this circumstance, identifying  $L'_i$  with  $L_i$  itself.<sup>4</sup> It is clear, then, that condition (3.2) is satisfied. Thus, for arbitrary mutually independent vector spaces  $L_1, \dots, L_k$  it is always possible to construct a space  $L$  containing all the  $L_i$  as subspaces that is their direct sum; that is,  $L = L_1 \oplus \dots \oplus L_k$ .

*Example 3.16* Let  $L_1$  be the vector space considered in Example 3.7, that is, the physical space that surrounds us, and let  $L_2 = \mathbb{R}$  be the real line, considered as the time axis. Operating as described above, we can define the direct sum  $L = L_1 \oplus L_2$ .

The vectors of the space  $L$  thus constructed are called *space–time events* and have the form  $(\mathbf{x}, t)$ , where  $\mathbf{x} \in L_1$  is the *space component*, and  $t \in L_2$  is the *time component*. For the addition of such vectors, the space components are added among themselves (as vectors in physical space, for example, according to the parallelogram law), while the time components are added to one another (as real numbers). Multiplication by a scalar is defined analogously. This space plays an important role in physics, in particular in the *theory of relativity*, where it is called *Minkowski space*. We remark that we still need to introduce some additional structure, namely a particular *quadratic form*. We shall return to this question in Sect. 7.7 (see p. 268).

## 3.2 Dimension and Basis

In this section we shall use the notion of *linear combination*, which in the case of a space of rows (or row space) of length  $n$  has already been introduced (see the

---

<sup>4</sup>More precisely, this identification is achieved with the help of the concept of *isomorphism* of vector spaces, which will be introduced below, in Sect. 3.5.



definition on p. 57). We shall now repeat that definition practically verbatim. In preparation, we observe that applying repeatedly the operations of vector addition and multiplication of a vector by a scalar, we can form more complex expressions, such as  $\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \cdots + \alpha_m \mathbf{x}_m$ , which, moreover, according to properties (a) and (b) of the definition of vector space, do not depend on the order of terms or the arrangement of parentheses (which is necessary in order that we be able to combine not only two vectors, but  $m$  of them).

**Definition 3.17** In the vector space  $L$ , let  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$  be  $m$  vectors. A vector  $\mathbf{y}$  is called a *linear combination* of these  $m$  vectors if

$$\mathbf{y} = \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \cdots + \alpha_m \mathbf{x}_m, \quad (3.4)$$

for some scalars  $\alpha_1, \alpha_2, \dots, \alpha_m$ .

The collection of all vectors that are linear combinations of some given vectors  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ , that is, those having the form (3.4) for all possible  $\alpha_1, \alpha_2, \dots, \alpha_m$ , clearly satisfies the definition of a subspace. This subspace is called the *linear span* of the vectors  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$  and is denoted by  $\langle \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m \rangle$ . It is clear that

$$\langle \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m \rangle = \langle \mathbf{x}_1 \rangle + \langle \mathbf{x}_2 \rangle + \cdots + \langle \mathbf{x}_m \rangle. \quad (3.5)$$

**Definition 3.18** Vectors  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$  are called *linearly dependent* if there exists a linear combination (3.4) equal to  $\mathbf{0}$  not all of whose coefficients  $\alpha_1, \alpha_2, \dots, \alpha_m$  are equal to zero. Otherwise,  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$  are said to be *linearly independent*.

Thus vectors  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$  are linearly dependent if for some scalars  $\alpha_1, \alpha_2, \dots, \alpha_m$ , one has

$$\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \cdots + \alpha_m \mathbf{x}_m = \mathbf{0}, \quad (3.6)$$

with at least one  $\alpha_i$  not equal to 0. For example, the vectors  $\mathbf{x}_1$  and  $\mathbf{x}_2 = -\mathbf{x}_1$  are linearly dependent. Conversely, the vectors  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$  are linearly independent if (3.6) holds only for  $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$ . In this case, the sum (3.5) is a direct sum, that is,

$$\langle \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m \rangle = \langle \mathbf{x}_1 \rangle \oplus \langle \mathbf{x}_2 \rangle \oplus \cdots \oplus \langle \mathbf{x}_m \rangle.$$

Here is a useful reformulation: Vectors  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$  are linearly dependent if and only if one of them is a linear combination of the others. Indeed, if

$$\mathbf{x}_i = \alpha_1 \mathbf{x}_1 + \cdots + \alpha_{i-1} \mathbf{x}_{i-1} + \alpha_{i+1} \mathbf{x}_{i+1} + \cdots + \alpha_m \mathbf{x}_m, \quad (3.7)$$

then we have the relationship (3.6) with  $\alpha_i = -1$ . Conversely, if in (3.6), the coefficient  $\alpha_i$  is not equal to 0, then if we transfer the term  $\alpha_i \mathbf{x}_i$  to the right-hand side and multiply both sides of the equality by the scalar  $-\alpha_i^{-1}$ , we obtain a representation of  $\mathbf{x}_i$  as a linear combination  $\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_m$ .

We are finally in a position to formulate the main definition of this section (and perhaps of the entire chapter).

**Definition 3.19** The *dimension* of a vector space  $L$  is the largest number of linearly independent vectors in the space, if such a number exists. The dimension of a vector space is denoted by  $\dim L$ , and if the greatest number of linearly independent vectors is finite, the space  $L$  is said to be *finite-dimensional*. If there is no maximum number of linearly independent vectors in  $L$ , then the space is said to be *infinite-dimensional*. The dimension of the vector space  $(\mathbf{0})$  is by definition equal to zero.

Thus the dimension of a vector space is equal to the natural number  $n$  if the space contains  $n$  linearly independent vectors and every set of  $m$  vectors for  $m > n$  is linearly dependent. A vector space is infinite-dimensional if there is a collection of  $n$  linearly independent vectors for every natural number  $n$ . Employing standard terminology, we shall call a space of dimension 1 a *line* and a space of dimension 2 a *plane*.

*Example 3.20* It is well known from elementary geometry (or from a course in analytic geometry) that vectors on a line, in the plane, or in the physical space that surrounds us form vector spaces of dimension 1, 2, and 3. This is the principal intuitive basis of the general definition of dimensionality.

*Example 3.21* The space of all polynomials in the variable  $t$  is clearly infinite-dimensional, since for an arbitrary number  $n$ , the polynomials  $1, t, t^2, \dots, t^{n-1}$  are linearly independent. The space of all continuous functions on the interval  $[a, b]$  is a fortiori infinite-dimensional.

The dimension of a vector space  $L$  depends not only on the set itself whose elements are the vectors of  $L$ , but also on the field over which it is defined. This will be made clear in the following examples.

*Example 3.22* Let  $L_1$  be the space whose vectors are the complex numbers, defined over the field  $\mathbb{C}$ . The operations of vector addition and multiplication by a scalar will be defined as the usual operations of addition and multiplication of complex numbers. Then it is easily seen from the definition that  $\dim L_1 = 1$ . If we now consider the vector space  $L_2$  likewise consisting of the complex numbers, but defined over the field  $\mathbb{R}$ , then we obtain  $\dim L_2 = 2$ . This, as we shall see, follows from the fact that every complex number is uniquely defined by a pair of real numbers (its real and imaginary parts). The frequently encountered expression “complex plane” implies the two-dimensional space  $L_2$  over the field  $\mathbb{R}$ , while the expression “complex line” indicates the one-dimensional space  $L_1$  over the field  $\mathbb{C}$ .

*Example 3.23* Let  $L$  be the vector space consisting of the real numbers, but defined over the field  $\mathbb{Q}$  of rational numbers (it is easy to see that all the conditions for the definition of a vector space are satisfied). In this case, in a linear combination (3.4), vectors  $x_i$  and  $y$  are real numbers, while  $\alpha_i$  is a rational number. By properties of sets of numbers proved in a course in real analysis, it follows that the space  $L$  is infinite-dimensional. Indeed, if the dimension of  $L$  were some finite number  $n$ , then

as we shall prove below, it would imply that there exist numbers  $x_1, \dots, x_n \in \mathbb{R}$  such that an arbitrary  $y \in \mathbb{R}$  could be written as a linear combination (3.4) with suitable coefficients  $\alpha_1, \dots, \alpha_n$  from the field  $\mathbb{Q}$ . But that would imply that the set of real numbers is countable, which, as is known from real analysis, is not the case.

It is obvious that the dimension of a subspace  $L'$  of a vector space  $L$  cannot be greater than the dimension of the entire space  $L$ .

**Theorem 3.24** *If the dimension of a subspace  $L'$  of a vector space  $L$  is equal to the dimension of  $L$ , then the subspace  $L'$  is equal to all of  $L$ .*

*Proof* Suppose  $\dim L' = \dim L = n$ . Then in  $L'$  one could find  $n$  linearly independent vectors  $x_1, \dots, x_n$ . If  $L' \neq L$ , then in  $L$  there would be some vector  $x \notin L'$ . Since  $\dim L = n$ , it follows that any  $n + 1$  vectors in this space are linearly dependent. In particular, the vectors  $x_1, \dots, x_n, x$  are linearly dependent. That is, there is a relationship

$$\alpha_1 x_1 + \dots + \alpha_n x_n + \alpha x = 0$$

with not all coefficients equal to zero. If we had  $\alpha = 0$ , then this would yield the linear dependence of the vectors  $x_1, \dots, x_n$ , which are linearly independent by assumption. This means that  $\alpha \neq 0$  and  $x = \beta_1 x_1 + \dots + \beta_n x_n$ ,  $\beta_i = -\alpha^{-1} \alpha_i$ , from which it follows that  $x$  is a linear combination of the vectors  $x_1, \dots, x_n$ . It clearly follows from the definition of a subspace that a linear combination of vectors in  $L'$  is itself a vector in  $L'$ . Hence we have  $x \in L'$ , and  $L' = L$ .  $\square$

If the dimension of a vector space  $L$  is finite,  $\dim L = n$ , and a subspace  $L' \subset L$  has dimension  $n - 1$ , then  $L'$  is called a *hyperplane* in  $L$ .

There is a defect in the definition of dimension given above: it is not effective. Theoretically, in order to determine the dimension of a vector space, it would be necessary to look at all systems of vectors  $x_1, \dots, x_m$  for various  $m$  in the space and determine whether each is linearly independent. With such a method, it is not so simple to determine the dimension of the row space of length  $n$  or of the space of polynomials of degree less than or equal to  $n$ . Therefore, we shall investigate the notion of dimension in greater detail.

**Definition 3.25** Vectors  $e_1, \dots, e_n$  of a vector space  $L$  are called a *basis* if they are linearly independent and every vector in the space  $L$  can be written as a linear combination of these vectors.

Thus if  $e_1, \dots, e_n$  is a basis of the space  $L$ , then for an arbitrary vector  $x \in L$  there exists an expression of the form

$$x = \alpha_1 e_1 + \alpha_2 e_2 + \dots + \alpha_n e_n. \quad (3.8)$$

**Theorem 3.26** *For an arbitrary vector  $x$ , the expression (3.8) is unique.*

*Proof* This is a direct consequence of the fact that the vectors  $\mathbf{e}_1, \dots, \mathbf{e}_n$  form a basis. Let us assume that there are two expressions

$$\mathbf{x} = \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \dots + \alpha_n \mathbf{e}_n, \quad \mathbf{x} = \beta_1 \mathbf{e}_1 + \beta_2 \mathbf{e}_2 + \dots + \beta_n \mathbf{e}_n.$$

Subtracting one equality from the other, we obtain

$$(\alpha_1 - \beta_1)\mathbf{e}_1 + (\alpha_2 - \beta_2)\mathbf{e}_2 + \dots + (\alpha_n - \beta_n)\mathbf{e}_n = \mathbf{0}.$$

But since the vectors  $\mathbf{e}_1, \dots, \mathbf{e}_n$  form a basis, then by definition, they are linearly independent. From this it follows that  $\alpha_1 = \beta_1, \alpha_2 = \beta_2, \dots, \alpha_n = \beta_n$ , as was to be proved.  $\square$

**Corollary 3.27** *If  $\mathbf{e}_1, \dots, \mathbf{e}_n$  is a basis of the vector space  $L$ , then  $L$  can be written in the form*

$$L = \langle \mathbf{e}_1 \rangle \oplus \langle \mathbf{e}_2 \rangle \oplus \dots \oplus \langle \mathbf{e}_n \rangle.$$

**Definition 3.28** The numbers  $\alpha_1, \dots, \alpha_n$  in the expression (3.8) are called the *coordinates* of the vector  $\mathbf{x}$  with respect to the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  (or *coordinates in that basis*).

*Example 3.29* An arbitrary vector  $\mathbf{e} \neq \mathbf{0}$  on a line (that is, a one-dimensional vector space) forms a basis of the line. For an arbitrary vector  $\mathbf{x}$  on the same line, we have the expression (3.8), which in the given case takes the form  $\mathbf{x} = \alpha \mathbf{e}$  with some scalar  $\alpha$ . This  $\alpha$  is the coordinate (in this case the only one) of the vector  $\mathbf{x}$  in the basis  $\mathbf{e}$ . If  $\mathbf{e}' \neq \mathbf{0}$  is another vector on the same line, then it provides another basis. We have seen that  $\mathbf{e}' = c\mathbf{e}$  for some scalar  $c \neq 0$  (since  $\mathbf{e}' \neq \mathbf{0}$ ). Therefore, from the relationship  $\mathbf{x} = \alpha \mathbf{e}$  we obtain that  $\mathbf{x} = \alpha c^{-1} \mathbf{e}'$ . Thus in the basis  $\mathbf{e}'$ , the coordinate of the vector  $\mathbf{x}$  is equal to  $\alpha c^{-1}$ .

Thus we have seen that the coordinates of a vector  $\mathbf{x}$  depend not only on the vector itself, but on the basis that we use (in the general case,  $\mathbf{e}_1, \dots, \mathbf{e}_n$ ). Consequently, the coordinates of a vector are not an “intrinsic geometric” property. The situation here is similar to the measurement of physical quantities: the length of a line segment or the mass of a body. Neither the one nor the other can be characterized by a number. It is necessary as well to have a unit of measurement: in the first case, the meter, centimeter, etc.; in the second, the kilogram, gram, etc. We shall encounter such a phenomenon repeatedly: some object (such as, for example, a vector) cannot be defined “in and of itself” by some set or other of numbers; rather, something similar to a unit of measurement (in our case, a basis) must be chosen. Here, there are always two possible points of view: either to choose some method of associating numbers with the object or to limit oneself to the study of its “purely intrinsic” properties, independent of the method of association. For example, in physics, we are interested in physical quantities themselves, but the laws of nature are usually expressed in the form of mathematical relationships among the numbers that characterize them. We will try to reconcile both points of view after defining how the

numbers that characterize the object change under different methods of associating numbers with the object. In particular, in Sect. 3.4, we shall consider the question of how the coordinates of a vector change under a change of basis.

In terms of the coordinates of vectors (relative to an arbitrary basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ ), it is easy to express the operations that enter into the definition of a vector space, namely the addition of vectors and the multiplication of a vector by a scalar. Namely, if  $\mathbf{x}$  and  $\mathbf{y}$  are two vectors, and

$$\mathbf{x} = \alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n, \quad \mathbf{y} = \beta_1 \mathbf{e}_1 + \dots + \beta_n \mathbf{e}_n,$$

then

$$\begin{aligned} \mathbf{x} + \mathbf{y} &= (\alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n) + (\beta_1 \mathbf{e}_1 + \dots + \beta_n \mathbf{e}_n) \\ &= (\alpha_1 + \beta_1) \mathbf{e}_1 + \dots + (\alpha_n + \beta_n) \mathbf{e}_n, \end{aligned} \quad (3.9)$$

and for an arbitrary scalar  $\alpha$ ,

$$\alpha \mathbf{x} = \alpha(\alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n) = (\alpha \alpha_1) \mathbf{e}_1 + \dots + (\alpha \alpha_n) \mathbf{e}_n, \quad (3.10)$$

so that the coordinates of vectors under addition are added, and under multiplication by a scalar, they are multiplied by that scalar.

It follows from the definition of a basis that if  $\dim L = n$  and  $\mathbf{e}_1, \dots, \mathbf{e}_n$  is any set of  $n$  linearly independent vectors in  $L$ , then they form a basis of  $L$ . Indeed, it suffices to verify that an arbitrary vector  $\mathbf{x} \in L$  can be written as a linear combination of these vectors. But from the definition of dimension,  $n + 1$  vectors  $\mathbf{x}, \mathbf{e}_1, \dots, \mathbf{e}_n$  are linearly dependent, that is,

$$\beta \mathbf{x} + \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \dots + \alpha_n \mathbf{e}_n = \mathbf{0}$$

for some scalars  $\beta, \alpha_1, \alpha_2, \dots, \alpha_n$ . In this case,  $\beta \neq 0$ , for otherwise, this would contradict the linear independence of the vectors forming the basis. But then

$$\mathbf{x} = -\beta^{-1} \alpha_1 \mathbf{e}_1 - \beta^{-1} \alpha_2 \mathbf{e}_2 - \dots - \beta^{-1} \alpha_n \mathbf{e}_n,$$

which was to be proved.

From the definition, it follows that if the dimension of a vector space  $L$  is equal to  $n$ , then there exist  $n$  linearly independent vectors in  $L$ , which by what we have proved, form a basis. Now we shall establish a more general fact.

**Theorem 3.30** *If  $\mathbf{e}_1, \dots, \mathbf{e}_m$  are linearly independent vectors in a vector space  $L$  of finite dimension  $n$ , then this set of vectors can be extended to a basis of  $L$ , that is, there exist vectors  $\mathbf{e}_i, m < i \leq n$ , such that  $\mathbf{e}_1, \dots, \mathbf{e}_m, \mathbf{e}_{m+1}, \dots, \mathbf{e}_n$  is a basis of  $L$ .*

*Proof* If the vectors  $\mathbf{e}_1, \dots, \mathbf{e}_m$  already form a basis, then  $m = n$ , and the theorem is proved. If they do not form a basis, then clearly  $m < n$ , and there exists a vector  $\mathbf{e}_{m+1}$  in  $L$  that is not a linear combination of  $\mathbf{e}_1, \dots, \mathbf{e}_m$ . Thus the vectors









We shall show that they constitute a basis of the set  $F(M)$ . Indeed, for any function  $f \in F(M)$  we have the obvious equality

$$f(y) = \sum_{x \in M} f(x) \delta_x(y), \quad (3.15)$$

from which it follows that an arbitrary function in the space  $F(M)$  can be expressed as a linear combination of the  $\delta_x$ ,  $x \in M$ . It is clear that the set of all delta functions is linearly independent, that is, they form a basis of the vector space  $F(M)$ . Since the number of functions in this collection is equal to the number of elements of the set  $M$ , the set  $F(M)$  is finite-dimensional, and  $\dim F(M)$  is equal to the number of elements in  $M$ . In the case that  $M = \mathbb{N}_n$  (see the definition on p. xi), then any function  $f \in F(\mathbb{N}_n)$  is uniquely determined by its values  $f(1), \dots, f(n)$ , which are its coordinates in the decomposition (3.15) with respect to the basis  $\delta_x$ ,  $x \in M$ . If we set  $a_i = f(i)$ , then the numbers  $(a_1, \dots, a_n)$  form a row, and this shows that the vector space  $F(\mathbb{N}_n)$  coincides with the space  $\mathbb{K}^n$ . In particular, the basis of the space  $F(\mathbb{N}_n)$  consisting of the delta functions coincides with the basis (3.14) of the space  $\mathbb{K}^n$ .

In many cases, Theorem 3.33 provides a simple method for finding the dimension of a vector space.

**Theorem 3.37** *The dimension of a vector space  $\langle \mathbf{x}_1, \dots, \mathbf{x}_m \rangle$  is equal to the maximal number of linearly independent vectors among the vectors  $\mathbf{x}_1, \dots, \mathbf{x}_m$ .*

Therefore, even though the definition of dimension requires the consideration of all the vectors in the space  $\langle \mathbf{x}_1, \dots, \mathbf{x}_m \rangle$ , Theorem 3.37 makes it possible to limit consideration to only the vectors  $\mathbf{x}_1, \dots, \mathbf{x}_m$ .

*Proof of Theorem 3.37* Let us set  $L' = \langle \mathbf{x}_1, \dots, \mathbf{x}_m \rangle$  and define by  $l$  the maximum number of linearly independent vectors among  $\mathbf{x}_1, \dots, \mathbf{x}_m$ . Changing the numeration if necessary, we may suppose that the first  $l$  vectors  $\mathbf{x}_1, \dots, \mathbf{x}_l$  are linearly independent. Let  $L'' = \langle \mathbf{x}_1, \dots, \mathbf{x}_l \rangle$ . It is clear that  $\mathbf{x}_1, \dots, \mathbf{x}_l$  form a basis of the space  $L''$ , and by Theorem 3.33,  $\dim L'' = l$ . We shall prove that  $L'' = L'$ , which will give us the result of Theorem 3.37. If  $l = m$ , then this is obvious. Suppose, then, that  $l < m$ . Then by our assumption, for any  $k = l + 1, \dots, m$ , the vectors  $\mathbf{x}_1, \dots, \mathbf{x}_l, \mathbf{x}_k$  are linearly dependent, that is, there is a linear combination  $\alpha_1 \mathbf{x}_1 + \dots + \alpha_l \mathbf{x}_l + \alpha_k \mathbf{x}_k = \mathbf{0}$  in which not all  $\alpha_i$  are equal to zero. And furthermore, it is necessary that  $\alpha_k \neq 0$ , since otherwise, we would obtain the linear dependence of the vectors  $\mathbf{x}_1, \dots, \mathbf{x}_l$ , which contradicts the hypothesis. Then

$$\mathbf{x}_k = -\alpha_k^{-1} \alpha_1 \mathbf{x}_1 - \alpha_k^{-1} \alpha_2 \mathbf{x}_2 - \dots - \alpha_k^{-1} \alpha_l \mathbf{x}_l,$$

that is, the vector  $\mathbf{x}_k$  is in  $L''$ . We have shown this for all  $k > l$ , but by construction, it is also true for  $k \leq l$ . This means that all vectors  $\mathbf{x}_k$  are in the space  $L''$ , and hence so are all linear combinations of them. Therefore, not only do we have  $L'' \subset L'$  (which is obvious by construction), but  $L' \subset L''$ , which shows that  $L'' = L'$ , as desired.  $\square$

**Theorem 3.38** *If  $L_1$  and  $L_2$  are two finite-dimensional vector spaces, then*

$$\dim(L_1 \oplus L_2) = \dim L_1 + \dim L_2.$$

*Proof* Let  $\dim L_1 = r$ ,  $\dim L_2 = s$ , let  $\mathbf{e}_1, \dots, \mathbf{e}_r$  be a basis of the space  $L_1$ , and let  $\mathbf{f}_1, \dots, \mathbf{f}_s$  be a basis of the space  $L_2$ . We shall show that the collection of  $r + s$  vectors  $\mathbf{e}_1, \dots, \mathbf{e}_r$ , and  $\mathbf{f}_1, \dots, \mathbf{f}_s$  forms a basis of the space  $L_1 \oplus L_2$ . By the definition of direct sum, every vector  $\mathbf{x} \in L_1 \oplus L_2$  can be expressed in the form  $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ , where  $\mathbf{x}_i \in L_i$ . But the vector  $\mathbf{x}_1$  is a linear combination of the vectors  $\mathbf{e}_1, \dots, \mathbf{e}_r$ , while the vector  $\mathbf{x}_2$  is a linear combination of the vectors  $\mathbf{f}_1, \dots, \mathbf{f}_s$ . As a result, we obtain a representation of the vector  $\mathbf{x}$  as a linear combination of the  $r + s$  vectors  $\mathbf{e}_1, \dots, \mathbf{e}_r, \mathbf{f}_1, \dots, \mathbf{f}_s$ . The linear independence of these vectors is just as easily verified. Suppose there exists a relationship

$$\alpha_1 \mathbf{e}_1 + \dots + \alpha_r \mathbf{e}_r + \beta_1 \mathbf{f}_1 + \dots + \beta_s \mathbf{f}_s = \mathbf{0}.$$

We set  $\mathbf{x}_1 = \alpha_1 \mathbf{e}_1 + \dots + \alpha_r \mathbf{e}_r$  and  $\mathbf{x}_2 = \beta_1 \mathbf{f}_1 + \dots + \beta_s \mathbf{f}_s$ . Then we have the equality  $\mathbf{x}_1 + \mathbf{x}_2 = \mathbf{0}$  with  $\mathbf{x}_i \in L_i$ . From this, by the definition of the direct sum, it follows that  $\mathbf{x}_1 = \mathbf{0}$  and  $\mathbf{x}_2 = \mathbf{0}$ . From the linear independence of the vectors  $\mathbf{e}_1, \dots, \mathbf{e}_r$ , it follows that  $\alpha_1 = 0, \dots, \alpha_r = 0$ , and similarly,  $\beta_1 = 0, \dots, \beta_s = 0$ .  $\square$

**Corollary 3.39** *For finite-dimensional spaces  $L_1, L_2, \dots, L_k$  for arbitrary  $k \geq 2$ , we have*

$$\dim(L_1 \oplus L_2 \oplus \dots \oplus L_k) = \dim L_1 + \dim L_2 + \dots + \dim L_k.$$

*Proof* The assertion follows readily from Theorem 3.38 by induction on  $k$ .  $\square$

**Corollary 3.40** *If  $L_1, \dots, L_r$  and  $L$  are vector spaces such that  $L = L_1 + \dots + L_r$ , and if  $\dim L = \dim L_1 + \dots + \dim L_r$ , then  $L = L_1 \oplus \dots \oplus L_r$ .*

*Proof* We select a basis in each of the  $L_i$  and combine them into a system of vectors  $\mathbf{e}_1, \dots, \mathbf{e}_n$ . By assumption, the number  $n$  of vectors in this system is equal to  $\dim L$ , and  $L = \langle \mathbf{e}_1, \dots, \mathbf{e}_n \rangle$ . By Theorem 3.37, the vectors  $\mathbf{e}_1, \dots, \mathbf{e}_n$  are linearly independent, and this implies that  $L = L_1 \oplus \dots \oplus L_r$ .  $\square$

These considerations make it possible to give a more visual, geometric, characterization of the notion of linear dependence. Namely, let us prove that vectors  $\mathbf{x}_1, \dots, \mathbf{x}_m$  are linearly dependent if and only if they are contained in a subspace  $L'$  of dimension less than  $m$ .

Indeed, let us denote by  $l$  the largest number of linearly independent vectors among  $\mathbf{x}_1, \dots, \mathbf{x}_m$ . Let us assume that these independent vectors are  $\mathbf{x}_1, \dots, \mathbf{x}_l$  and set  $L' = \langle \mathbf{x}_1, \dots, \mathbf{x}_l \rangle$ . Then for  $l = m$ , the vectors  $\mathbf{x}_1, \dots, \mathbf{x}_m$  are linearly independent, and our assertion follows from the definition of dimension. If  $l < m$ , then all the vectors  $\mathbf{x}_1, \dots, \mathbf{x}_m$  are contained in the subspace  $L'$ , whose dimension, by Theorem 3.33, is  $l$ , and the assertion is correct.

Using the concepts introduced thus far, it is possible to prove a useful generalization of Theorem 3.38.

**Theorem 3.41** *For any two finite-dimensional vector spaces  $L_1$  and  $L_2$ , one has the equality*

$$\dim(L_1 + L_2) = \dim L_1 + \dim L_2 - \dim(L_1 \cap L_2). \quad (3.16)$$

Theorem 3.38 is obtained as a simple corollary of Theorem 3.41. Indeed, if  $L_1 + L_2 = L_1 \oplus L_2$ , then by Corollary 3.15, the intersection  $L_1 \cap L_2$  is equal to  $(\mathbf{0})$ , and it remains only to use the fact that  $\dim(\mathbf{0}) = 0$ .

*Proof of Theorem 3.41* Let us set  $L_0 = L_1 \cap L_2$ . From Corollary 3.31, it follows that there exist subspaces  $L'_1 \subset L_1$  and  $L'_2 \subset L_2$  such that

$$L_1 = L_0 \oplus L'_1, \quad L_2 = L_0 \oplus L'_2. \quad (3.17)$$

Formula (3.16) follows easily from the equality  $L_1 + L_2 = L_0 \oplus L'_1 \oplus L'_2$ . Indeed, since  $L_0 = L_1 \cap L_2$ , then in view of relationship (3.17) and Theorem 3.38, we obtain  $L_1 + L_2 = L_1 \oplus L'_2$ , and therefore,

$$\dim(L_1 + L_2) = \dim L_1 + \dim L'_2 = \dim L_1 + \dim L_2 - \dim L_0,$$

which yields relationship (3.16).

Let us prove that  $L_1 + L_2 = L_0 \oplus L'_1 \oplus L'_2$ . It is clear that each subspace  $L_0, L'_1, L'_2$  is contained in  $L_1 + L_2$ , so that their sum  $L_0 + L'_1 + L'_2$  is also contained in  $L_1 + L_2$ . But an arbitrary vector  $z \in L_1 + L_2$  can be represented in the form  $z = x + y$ , where  $x \in L_1, y \in L_2$ , and in view of relationship (3.17), we have the representations  $x = u + v$  and  $y = u' + w$ , where  $u, u' \in L_0, v \in L'_1, w \in L'_2$ , from which we obtain  $z = x + y = (u + u') + v + w$ , and this means that the vector  $z$  is contained in  $L_0 + L'_1 + L'_2$ . From this, it follows that

$$L_1 + L_2 = L_0 + L'_1 + L'_2 = L_1 \oplus L'_2.$$

But  $L_1 \cap L'_2 = (\mathbf{0})$ , since the vector  $x \in L_1 \cap L'_2$  is contained both in  $L_1 \cap L_2 = L_0$  and in  $L'_2$ , while in view of (3.17), the intersection  $L_0 \cap L'_2$  is equal to  $(\mathbf{0})$ . As a result, we obtain the required equality

$$L_1 + L_2 = (L_0 \oplus L'_1) + L'_2 = (L_0 \oplus L'_1) \oplus L'_2 = L_0 \oplus L'_1 \oplus L'_2,$$

which, as we have seen, proves Theorem 3.41. □

**Corollary 3.42** *Let  $L_1$  and  $L_2$  be subspaces of a finite-dimensional vector space  $L$ . Then from the inequality  $\dim L_1 + \dim L_2 > \dim L$ , it follows that  $L_1 \cap L_2 \neq (\mathbf{0})$ , that is, the subspaces  $L_1$  and  $L_2$  have a nonzero vector in common.*

Indeed, in this case,  $L_1 + L_2 \subset L$ , which means that  $\dim(L_1 + L_2) \leq \dim L$ . Taking this into account, we obtain from (3.16) that

$$\dim(L_1 \cap L_2) = \dim L_1 + \dim L_2 - \dim(L_1 + L_2) \geq \dim L_1 + \dim L_2 - \dim L > 0,$$

from which it follows that  $L_1 \cap L_2 \neq \{0\}$ .

For example, two planes passing through the origin in three-dimensional space have a straight line in common.

We shall now obtain an expression for the dimension of a subspace  $\langle \mathbf{a}_1, \dots, \mathbf{a}_m \rangle$  using the theory of determinants. Let  $\mathbf{a}_1, \dots, \mathbf{a}_m$  be vectors in the space  $L$ , and let  $\mathbf{e}_1, \dots, \mathbf{e}_n$  be some basis of  $L$ . We shall write the coordinates of the vector  $\mathbf{a}_i$  in this basis as the  $i$ th row of a matrix  $A$ :

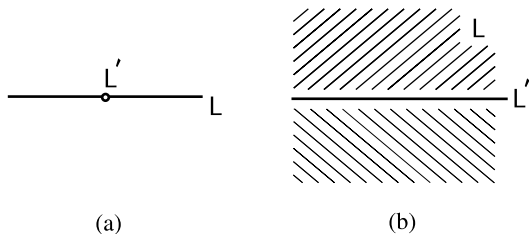
$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

**Theorem 3.43** *The dimension of the vector space  $\langle \mathbf{a}_1, \dots, \mathbf{a}_m \rangle$  is equal to the rank of the matrix  $A$ .*

*Proof* The linear dependence of the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_k$  for  $k \leq m$  is equivalent to the linear dependence of the rows of the matrix  $A$  consisting of the same numbers. In Theorem 2.41 we proved that if the rank of a matrix is equal to  $r$ , then all of its rows are linear combinations of some collection of  $r$  of its rows. From this it follows already that  $\dim\langle \mathbf{a}_1, \dots, \mathbf{a}_m \rangle \leq r$ . But in fact, from the proof of the same Theorem 2.41, it follows that for such a collection of  $r$  rows, one may take any  $r$  rows of the matrix in which there is a nonzero minor of order  $r$  (see the remark following Theorem 2.41). Let us show that such a collection of  $r$  rows is linearly independent, from which we will already have a proof of Theorem 3.43. We may assume that a nonzero minor  $M_r$  is located in the first  $r$  columns and first  $r$  rows of the matrix  $A$ . We then have to establish the linear independence of the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_r$ . If we assume that  $\alpha_1 \mathbf{a}_1 + \cdots + \alpha_r \mathbf{a}_r = \mathbf{0}$ , then if we focus attention on only the first  $r$  coordinates of the vectors, we obtain  $r$  homogeneous linear equations in the unknown coefficients  $\alpha_1, \dots, \alpha_r$ . It is easy to see that the determinant of the matrix of this system is equal to  $M_r \neq 0$ , and as a consequence, it has a unique solution, which is the zero solution:  $\alpha_1 = 0, \dots, \alpha_r = 0$ . That is, the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_r$  are indeed linearly independent.  $\square$

In the past, Theorem 3.43 was formulated in the following form, which is also sometimes useful. Consider the vector space  $\mathbb{K}^n$  of rows of length  $n$  (where  $\mathbb{K}$  is the field of real numbers, the field of complex numbers, or an arbitrary field). Then the vectors  $\mathbf{a}_i$  will be rows of length  $n$  (in our case, the rows of the matrix  $A$ ). From the proof of Theorem 3.43 we have at once the following corollary.

**Fig. 3.4** Hyperplanes in a vector space



**Corollary 3.44** *The rank of a matrix  $A$  is equal to the largest number of linearly independent rows of  $A$ .*

From this, we obtain the following unexpected result.

**Corollary 3.45** *The rank of a matrix  $A$  is also equal to the largest number of linearly independent columns of  $A$ .*

This follows at once from the definition of the rank of a matrix and Theorem 2.32.

To conclude this section, let us examine in greater detail the case of real vector spaces, and to this end, introduce some important notions that will be used in the sequel.

Let  $L'$  be a hyperplane in the finite-dimensional vector space  $L$ , that is,  $\dim L' = \dim L - 1$ . Then this hyperplane divides  $L$  into two parts, as shown in Fig. 3.4 for the case of a line and a plane.

Indeed, since  $L' \neq L$ , there exists a vector  $e \in L$ ,  $e \notin L'$ . From this, it follows that  $L = L' \oplus \langle e \rangle$ . For according to the choice of  $e$ , the intersection  $L' \cap \langle e \rangle$  is equal to  $\{0\}$ , and by Theorem 3.38, we have the equality

$$\dim(L' \oplus \langle e \rangle) = \dim L' + 1 = \dim L,$$

from which we obtain, with the help of Theorem 3.24, that  $L' \oplus \langle e \rangle = L$ . Thus an arbitrary vector  $x \in L$  can be uniquely expressed in the form

$$x = \alpha e + u, \quad u \in L', \quad (3.18)$$

where  $\alpha$  is some scalar. Since the scalars in our case are real, it makes sense to talk about their sign. The collection of vectors  $x$  expressed as in (3.18) for which  $\alpha > 0$  is denoted by  $L^+$ . Likewise, the set of vectors  $x$  of the form (3.18) for which  $\alpha < 0$  is denoted by  $L^-$ . The sets  $L^+$  and  $L^-$  are called *half-spaces* of the space  $L$ . Clearly,  $L \setminus L' = L^+ \cup L^-$ .

Of course, our construction depends not only on the hyperplane  $L'$ , but also on the choice of the vector  $e \notin L'$ . It is important to note that with a change in the vector  $e$ , the half-spaces  $L^+$  and  $L^-$  might change, but the pair  $(L^+, L^-)$  will remain as before; that is, either the spaces do not change at all, or else they exchange places. Indeed, let  $e' \notin L'$  be some other vector. Then it can be represented in the form  $e' = \lambda e + v$ ,

where the number  $\lambda$  is nonzero and  $\mathbf{v}$  is in  $L'$ . This means that  $\mathbf{e} = \lambda^{-1}(\mathbf{e}' - \mathbf{v})$ . Then for an arbitrary vector  $\mathbf{x}$  from (3.18), we obtain, as in (3.18), the representation

$$\mathbf{x} = \alpha\lambda^{-1}(\mathbf{e}' - \mathbf{v}) + \mathbf{u} = \alpha\lambda^{-1}\mathbf{e}' + \mathbf{u}', \quad \mathbf{u}' \in L',$$

where  $\mathbf{u}' = \mathbf{u} - \alpha\lambda^{-1}\mathbf{v}$ , and we see that in passing from  $\mathbf{e}$  to  $\mathbf{e}'$ , the scalar  $\alpha$  in the decomposition (3.18) is multiplied by  $\lambda^{-1}$ . Hence the half-spaces  $L^+$  and  $L^-$  do not change if  $\lambda > 0$ , and they exchange places if  $\lambda < 0$ .

The above definition of decomposition of a real vector space  $L$  by a hyperplane  $L'$  has a natural interpretation in topological terms (see pp. xvii–xix). Readers not interested in this aspect of these ideas can skip the following five paragraphs.

If we wish to use topological terminology, then we are going to have to introduce on  $L$  the notion of *convergence* of a sequence of vectors. We shall do this using the notion of a metric (see p. xviii). Let us choose in  $L$  an arbitrary basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , and for vectors  $\mathbf{x} = \alpha_1\mathbf{e}_1 + \dots + \alpha_n\mathbf{e}_n$  and  $\mathbf{y} = \beta_1\mathbf{e}_1 + \dots + \beta_n\mathbf{e}_n$ , we define the number  $r(\mathbf{x}, \mathbf{y})$  by means of formula

$$r(\mathbf{x}, \mathbf{y}) = |\alpha_1 - \beta_1| + \dots + |\alpha_n - \beta_n|.$$

It easily follows from the properties of absolute value that all three conditions in the definition of a metric space are satisfied. Thus the vector space  $L$  and all of its subspaces are metric spaces with the metric  $r(\mathbf{x}, \mathbf{y})$ , and for a sequence of vectors there is automatically defined the notion of convergence:  $\mathbf{x}_k \rightarrow \mathbf{x}$  as  $k \rightarrow \infty$  if  $r(\mathbf{x}_k, \mathbf{x}) \rightarrow 0$  as  $k \rightarrow \infty$ . In other words, if  $\mathbf{x} = \alpha_1\mathbf{e}_1 + \dots + \alpha_n\mathbf{e}_n$  and  $\mathbf{x}_k = \alpha_{1k}\mathbf{e}_1 + \dots + \alpha_{nk}\mathbf{e}_n$ , then the convergence  $\mathbf{x}_k \rightarrow \mathbf{x}$  is equivalent to the convergence of the  $n$  coordinate sequences:  $\alpha_{ik} \rightarrow \alpha_i$  for all  $i = 1, \dots, n$ . We observe that in the definition of  $r(\mathbf{x}, \mathbf{y})$ , we have used the coordinates of the vectors  $\mathbf{x}$  and  $\mathbf{y}$  in some basis, and consequently, the metric obtained depends on the choice of basis. Nevertheless, the notion of convergence does not depend on the choice of basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ . This follows easily from the formulas (3.35) relating the coordinates of a vector in various bases, which will be presented later.

The meaning of a partition  $L \setminus L' = L^+ \cup L^-$  consists in the fact that the metric space  $L \setminus L'$  is not path-connected, while  $L^+$  and  $L^-$  are its path-connected components.

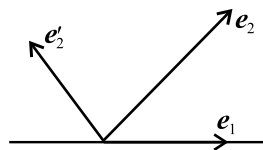
Indeed, let us suppose that in the metric space  $L \setminus L'$ , there exists a deformation of the vector  $\mathbf{x}$  to  $\mathbf{y}$ , that is, a continuous mapping  $\mathbf{f} : [0, 1] \rightarrow L \setminus L'$  such that  $\mathbf{f}(0) = \mathbf{x}$  and  $\mathbf{f}(1) = \mathbf{y}$ . Then by formula (3.18), we have the representation

$$\mathbf{x} = \alpha\mathbf{e} + \mathbf{u}, \quad \mathbf{y} = \beta\mathbf{e} + \mathbf{v}, \quad \mathbf{f}(t) = \gamma(t)\mathbf{e} + \mathbf{w}(t), \quad (3.19)$$

where  $\mathbf{u}, \mathbf{v} \in L'$  and  $\mathbf{w}(t) \in L'$  for all  $t \in [0, 1]$ , and  $\gamma(t)$  is a function taking real values, continuous in the interval  $[0, 1]$ , and moreover,  $\gamma(0) = \alpha$  and  $\gamma(1) = \beta$ .

If  $\mathbf{x} \in L^+$  and  $\mathbf{y} \in L^-$ , then  $\alpha > 0$  and  $\beta < 0$ , and by properties of continuous functions known from calculus,  $\gamma(\tau) = 0$  for some  $0 < \tau < 1$ . But then the vector  $\mathbf{f}(\tau) = \mathbf{w}(\tau)$  is contained within the hyperplane  $L'$ , and it follows that vectors  $\mathbf{x}$  and  $\mathbf{y}$  cannot be deformed into each other in the set  $L \setminus L'$ . Therefore, the metric space

**Fig. 3.5** Bases assigning one and the same flag



$L \setminus L'$  is not path-connected. But if  $x, y \in L^+$  or  $x, y \in L^-$ , then in the representations (3.19) for these vectors, the numbers  $\alpha$  and  $\beta$  have the same sign. Then, as is easily seen, the mapping  $f(t) = (1-t)x + ty$ ,  $t \in [0, 1]$ , determines a continuous deformation of  $x$  to  $y$  in the set  $L^+$  or  $L^-$ , respectively.

From these considerations, it is easy to obtain a proof of the previous assertion without using any formulas.

If we distinguish one of the two half-spaces  $L^+$  and  $L^-$  (we shall denote the half-space thus distinguished by  $L^+$ ), then the pair  $(L, L')$  is said to be *directed*. For example, in the case of a line (Fig. 3.4(a)), this corresponds to a choice of the direction of the line  $L$ .

Using these concepts, we can obtain a more visual idea of the notion of basis (in the case of a real vector space).

**Definition 3.46** A *flag* in a finite-dimensional vector space  $L$  is a sequence of subspaces

$$(0) \subset L_1 \subset L_2 \subset \cdots \subset L_n = L \quad (3.20)$$

such that

- (a)  $\dim L_i = i$  for all  $i = 1, \dots, n$ .
- (b) Each pair  $(L_i, L_{i-1})$  is directed.

It is clear that in view of condition (a), the subspace  $L_{i-1}$  is a hyperplane in  $L_i$ , and therefore the above definition of directedness is applicable.

Every basis  $e_1, \dots, e_n$  of a space  $L$  determines a particular flag. Namely, we set  $L_i = \langle e_1, \dots, e_i \rangle$ , and to apply directedness to the pair  $(L_i, L_{i-1})$ , we select in the collection of half-spaces  $L_i^+$  the one determined by the vector  $e_i$  (clearly,  $e_i \notin L_{i-1}$ ).

However, we must observe that different bases of the space  $L$  can determine one and the same flag. For example, in Fig. 3.5, the bases  $(e_1, e_2)$  and  $(e_1, e'_2)$  determine the same flag in the plane. But later, in Sect. 7.2, we shall meet a situation in which there is defined a bijection between the bases of a vector space and its flags (this is accomplished through the selection of some special bases).

### 3.3 Linear Transformations of Vector Spaces

Here we shall present a very broad generalization of the notion of linear function, with which our course began. The generalization occurs in two aspects. First, in Sect. 1.1, a linear function was defined as a function of rows of length  $n$ . Here, we

shall replace the rows of given length with vectors of an arbitrary vector space  $L$ . Second, the value of the linear function in Sect. 1.1 was considered a number, that is, in other words, an element of the space  $\mathbb{R}^1$  or  $\mathbb{C}^1$  or  $\mathbb{K}^1$  for an arbitrary field  $\mathbb{K}$ . We shall now replace the numbers with vectors in an arbitrary vector space  $M$ . Thus our definition will include two vector spaces  $L$  and  $M$ . The reader may consider both spaces real, complex, or defined over an arbitrary field  $\mathbb{K}$ , but it must be the same field for both  $L$  and  $M$ . In this case, we shall speak about the elements of the field using the same conventions that we established in Sect. 3.1 for scalars (see p. 82).

Let us recall that a linear function is defined by properties (1.8) and (1.9), presented in Theorem 1.3 on page 3. The following definition is analogous to this.

**Definition 3.47** A *linear transformation* of a vector space  $L$  to another vector space  $M$  is a mapping  $\mathcal{A} : L \rightarrow M$  that assigns to each vector  $\mathbf{x} \in L$  some vector  $\mathcal{A}(\mathbf{x}) \in M$  and exhibits the following properties:

$$\begin{aligned}\mathcal{A}(\mathbf{x} + \mathbf{y}) &= \mathcal{A}(\mathbf{x}) + \mathcal{A}(\mathbf{y}), \\ \mathcal{A}(\alpha\mathbf{x}) &= \alpha\mathcal{A}(\mathbf{x})\end{aligned}\tag{3.21}$$

for every scalar  $\alpha$  and all vectors  $\mathbf{x}$  and  $\mathbf{y}$  in the space  $L$ .

A linear transformation is also called an *operator* or (only in the case that  $M = L$ ) an *endomorphism*.

Let us note one obvious but useful property that follows directly from the definitions.

**Proposition 3.48** *Under any linear transformation, the image of the null vector is the null vector. More precisely, since we may be dealing with two different vector spaces, we might reformulate the statement in the following form: if  $\mathcal{A} : L \rightarrow M$  is a linear transformation, and  $\mathbf{0} \in L$  and  $\mathbf{0}' \in M$  are the null vectors in the vector spaces  $L$  and  $M$ , then  $\mathcal{A}(\mathbf{0}) = \mathbf{0}'$ .*

*Proof* By the definition of a vector space, for an arbitrary vector  $\mathbf{x} \in L$ , there exists an additive inverse  $-\mathbf{x} \in L$ , that is, a vector such that  $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$ , and moreover (see p. 82), the vector  $-\mathbf{x}$  is obtained by multiplying  $\mathbf{x}$  by the number  $-1$ . Applying the linear transformation  $\mathcal{A}$  to both sides of the equality  $\mathbf{0} = \mathbf{x} + (-\mathbf{x})$ , then in view of properties (3.21), we obtain  $\mathcal{A}(\mathbf{0}) = \mathcal{A}(\mathbf{x}) + \mathcal{A}(-\mathbf{x}) = \mathcal{A}(\mathbf{x}) - \mathcal{A}(\mathbf{x}) = \mathbf{0}'$ , since for the vector  $\mathcal{A}(\mathbf{x})$  of the space  $M$ , the vector  $-\mathcal{A}(\mathbf{x})$  is its additive inverse, and their sum is  $\mathbf{0}'$ .  $\square$

**Example 3.49** For an arbitrary vector space  $L$ , the *identity mapping* defines a linear transformation  $\mathcal{E}(\mathbf{x}) = \mathbf{x}$ , for every  $\mathbf{x} \in L$ , from the space  $L$  to itself.

**Example 3.50** A rotation of the plane  $\mathbb{R}^2$  through some angle about the origin is a linear transformation (here  $L = M = \mathbb{R}^2$ ). The conditions of (3.21) are clearly satisfied here.



*Example 3.51* If  $L$  is the space of continuously differentiable functions on an interval  $[a, b]$ , and  $M$  is the space of continuous functions on the same interval, and if for  $\mathbf{x} = f(t)$ , we define  $\mathcal{A}(\mathbf{x}) = f'(t)$ , then the mapping  $\mathcal{A} : L \rightarrow M$  is a linear transformation.

*Example 3.52* If  $L$  is the space of twice continuously differentiable functions on an interval  $[a, b]$ ,  $M$  is the same space as in the previous example,  $q(t)$  is some continuous function on the interval  $[a, b]$ , and for  $\mathbf{x} = f(t)$  we set  $\mathcal{A}(\mathbf{x}) = f''(t) + q(t)f(t)$ , then the mapping  $\mathcal{A} : L \rightarrow M$  is a linear transformation. In analysis, it is known as the *Sturm–Liouville operator*.

*Example 3.53* Let  $L$  be the space of all polynomials, and for  $\mathbf{x} = f(t)$ , as in Example 3.51, we set  $\mathcal{A}(\mathbf{x}) = f'(t)$ . Clearly,  $\mathcal{A} : L \rightarrow L$  is a linear transformation (that is, here we have  $M = L$ ). But if  $L$  is the space of polynomials of degree at most  $n$ , and  $M$  is the space of polynomials of degree at most  $n - 1$ , then the same formula gives a linear transformation  $\mathcal{A} : L \rightarrow M$ .

*Example 3.54* Suppose we are given the representation of a space  $L$  as a direct sum of two subspaces:  $L = L' \oplus L''$ . This means that every vector  $\mathbf{x} \in L$  can be uniquely represented in the form  $\mathbf{x} = \mathbf{x}' + \mathbf{x}''$ , where  $\mathbf{x}' \in L'$  and  $\mathbf{x}'' \in L''$ . Assigning to each vector  $\mathbf{x} \in L$  the term  $\mathbf{x}' \in L'$  in this representation gives a mapping  $\mathcal{P} : L \rightarrow L'$ ,  $\mathcal{P}(\mathbf{x}) = \mathbf{x}'$ . A simple verification shows that  $\mathcal{P}$  is a linear transformation. It is called the *projection* onto the subspace  $L'$  parallel to  $L''$ . In this case, for the vector  $\mathbf{x} \in L$ , its image  $\mathcal{P}(\mathbf{x}) \in L'$  is called the *projection* vector of  $\mathbf{x}$  onto  $L'$  parallel to  $L''$ . Analogously, for any subset  $X \subset L$ , its image  $\mathcal{P}(X) \subset L'$  is called the *projection* of  $X$  onto  $L'$  parallel to  $L''$ .

*Example 3.55* Let  $L = M$  and  $\dim L = \dim M = 1$ . Then  $L = M = \langle \mathbf{e} \rangle$ , where  $\mathbf{e}$  is some nonnull vector and  $\mathcal{A}(\mathbf{e}) = \alpha \mathbf{e}$ , where  $\alpha$  is a scalar. From the definition of a linear transformation, it follows directly that  $\mathcal{A}(\mathbf{x}) = \alpha \mathbf{x}$  for every vector  $\mathbf{x} \in L$ . Consequently, such is the general form of all linear transformations  $\mathcal{A} : L \rightarrow L$  in the case  $\dim L = 1$ .

In the sequel, we shall consider the case that the dimensions of the spaces  $L$  and  $M$  are finite. This means that in  $L$ , there exists some basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , and in  $M$ , there is a basis  $\mathbf{f}_1, \dots, \mathbf{f}_m$ . Then every vector  $\mathbf{x} \in L$  can be written in the form

$$\mathbf{x} = \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \dots + \alpha_n \mathbf{e}_n.$$

Using the relationship (3.21) several times, we shall obtain that for any linear transformation  $\mathcal{A} : L \rightarrow M$ , the image of the vector  $\mathbf{x}$  is equal to

$$\mathcal{A}(\mathbf{x}) = \alpha_1 \mathcal{A}(\mathbf{e}_1) + \alpha_2 \mathcal{A}(\mathbf{e}_2) + \dots + \alpha_n \mathcal{A}(\mathbf{e}_n). \quad (3.22)$$



**Definition 3.56** The matrix  $A$  in (3.26) is called the *matrix of the linear transformation*  $\mathcal{A} : L \rightarrow M$  given by formula (3.23) in the bases  $e_1, \dots, e_n$  and  $f_1, \dots, f_m$ .

In other words, the matrix  $A$  of the linear transformation  $\mathcal{A}$  is a matrix whose  $i$ th column consists of the coordinates of the vector  $\mathcal{A}(e_i)$  in the basis  $f_1, \dots, f_m$ . We would like to emphasize that the coordinates are written in the columns, and not in the rows (which, of course, also would have been possible), which has a number of advantages. It is clear that the matrix of the linear transformation depends on both bases  $e_1, \dots, e_n$  and  $f_1, \dots, f_m$ . The situation here is the same as with the coordinates of a vector. A linear transformation has no matrix “in and of itself”: in order to associate a matrix with the transformation, it is necessary to choose bases in the spaces  $L$  and  $M$ .

Using matrix multiplication, as defined in Sect. 2.9, one can write formula (3.25) in a more compact form. To do so, we introduce the following notation: Let  $\alpha$  be a row vector (a matrix of type  $(1, n)$ ), with coordinates  $\alpha_1, \dots, \alpha_n$ , and let  $\beta$  be a row vector with coordinates  $\beta_1, \dots, \beta_n$ . Similarly, let  $[\alpha]$  be a column vector (a matrix of type  $(n, 1)$ ), consisting of the same coordinates  $\alpha_1, \dots, \alpha_n$ , only now written vertically, and let  $[\beta]$  be a column vector consisting of  $\beta_1, \dots, \beta_n$ , that is,

$$[\alpha] = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}, \quad [\beta] = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}.$$

It is clear that  $\alpha$  and  $[\alpha]$  are interchanged under the transpose operation, that is,  $\alpha^* = [\alpha]$ , and similarly,  $\beta^* = [\beta]$ . Recalling the definition of matrix multiplication, we see that formula (3.25) has the form

$$[\beta] = A[\alpha] \quad \text{or} \quad \beta = \alpha A^*. \quad (3.27)$$

The formulas that we have obtained show that with the chosen bases, a linear transformation is uniquely determined by its matrix. Conversely, having chosen bases for the vector spaces  $L$  and  $M$  in some way, then if we define the mapping  $\mathcal{A} : L \rightarrow M$  with the help of relationships (3.22) and (3.23) with arbitrary matrix  $A = (a_{ij})$ , it is easy to verify that  $\mathcal{A}$  will be a linear transformation. Therefore, there exists a bijection between the set  $\mathcal{L}(L, M)$  of linear transformations  $L$  into  $M$  and the set of matrices of type  $(n, m)$ . It is the choice of bases in the spaces  $L$  and  $M$  that determines this correspondence. In the following section, we shall explain precisely how the matrix of a linear transformation depends on the choice of bases.

We shall denote the space of all linear transformations of the space  $L$  into  $M$  by  $\mathcal{L}(L, M)$ . This set can itself be viewed as a vector space if for the mappings  $\mathcal{A}$  and  $\mathcal{B}$  in  $\mathcal{L}(L, M)$  we define the vector sum and the product of a vector and a scalar  $\alpha$  by the following formulas:

$$\begin{aligned} (\mathcal{A} + \mathcal{B})(x) &= \mathcal{A}(x) + \mathcal{B}(x), \\ (\alpha \mathcal{A})(x) &= \alpha \mathcal{A}(x). \end{aligned} \quad (3.28)$$



Formulas (3.23) and (3.30) represent two linear replacements in which the vectors play the role of the variables, whereas in other respects, they are no different from linear replacements of variables as examined by us earlier (see p. 62). Consequently, the result of sequentially applying these replacements will be the same as in Sect. 2.9, namely linear replacement with the matrix  $BA$ ; that is, we obtain the relationship

$$(\mathcal{B}\mathcal{A})(\mathbf{e}_i) = \sum_{j=1}^l c_{ij} \mathbf{g}_j, \quad i = 1, \dots, n,$$

where the matrix  $C = (c_{ij})$  of the transformation  $\mathcal{B}\mathcal{A}$  is  $BA$ . We have thus established that *the composition of linear transformations corresponds to the multiplication of their matrices, taken in the same order.*

We observe that we have thus obtained a shorter and more natural proof of the associativity of matrix multiplication (formula (2.52)) in Sect. 2.9. Indeed, the associativity of the composition of arbitrary mappings of sets is well known (p. xiv), and in view of the established connection between linear transformations and their matrices (in whatever selected bases), we obtain the associativity of the matrix product.

The operations of addition and composition of linear transformations are connected by the relationships

$$\mathcal{A}(\mathcal{B} + \mathcal{C}) = \mathcal{A}\mathcal{B} + \mathcal{A}\mathcal{C}, \quad (\mathcal{A} + \mathcal{B})\mathcal{C} = \mathcal{A}\mathcal{C} + \mathcal{B}\mathcal{C},$$

called the *distributive property*. To prove this, one may either use the definitions of addition and composition defined above together with the well-known property of the distributivity of the real and complex numbers (or the elements of any set  $\mathbb{K}$ , since it derives from the properties of a field) or derive the distributivity of linear transformations from what was proved in Sect. 2.9 regarding distributivity of addition and multiplication of matrices (formula (2.53)), again using the connection established above between a linear transformation and its matrix.

### 3.4 Change of Coordinates

We have seen that the coordinates of a vector relative to a basis depend on which basis in the vector space we have chosen. We have seen as well that the matrix of a linear transformation of vector spaces depends on the choice of bases in both vector spaces. We shall now establish an explicit form of this dependence both for vectors and for transformations.

Let  $\mathbf{e}_1, \dots, \mathbf{e}_n$  be a certain basis of the vector space  $L$ . By Corollary 3.34, a basis of the given vector space consists of a fixed number of vectors, equal to  $\dim L$ . Let  $\mathbf{e}'_1, \dots, \mathbf{e}'_n$  be another basis of  $L$ . By definition, every vector  $\mathbf{x} \in L$  is a linear combination of the vectors  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , that is, it can be expressed in the form

$$\mathbf{x} = \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \dots + \alpha_n \mathbf{e}_n \quad (3.31)$$



Relationships (3.35) are called *change-of-coordinate formulas* for a vector. Such a formula represents a linear change of variables, with the help of the matrix  $C$  consisting of the coefficients  $c_{ij}$ , but in an order different from that in (3.33). In particular,  $C$  is the transpose of the matrix of coefficients (3.33). The matrix  $C$  is called the *transition matrix* from the basis  $\mathbf{e}'_1, \dots, \mathbf{e}'_n$  to the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , since with its help, the coordinates of a vector in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  are expressed in terms of its coordinates in the basis  $\mathbf{e}'_1, \dots, \mathbf{e}'_n$ .

Using the product rule for matrices, the formula for the change of coordinates can be written in a more compact form. To this end, we shall use notation from the preceding section:  $\alpha$  is a row vector consisting of the coordinates  $\alpha_1, \dots, \alpha_n$ , and  $[\alpha]$  is a column vector consisting of the very same coordinates. Keeping in mind the definition of matrix multiplication (Sect. 2.9), we see that formula (3.35) takes the form

$$[\alpha] = C[\alpha'] \quad \text{or} \quad \alpha = \alpha' C^*. \quad (3.36)$$

*Remark 3.57* It is not difficult to see that the formulas for changing coordinates are quite similar to the formulas for a linear transformation. More precisely, relationships (3.35) and (3.36) are special cases of (3.25) and (3.27) for  $m = n$ , for example, if the vector space  $\mathbf{M}$  coincides with  $\mathbf{L}$ . This allows an interpretation of changing coordinates (that is, changing bases) of a vector space  $\mathbf{L}$  as a linear transformation  $\mathcal{A} : \mathbf{L} \rightarrow \mathbf{L}$ .

Similarly, if we substitute expressions (3.34) for vectors  $\mathbf{e}_i$  into (3.31), we obtain the relationship

[illegible]

similar to (3.35). Formula (3.37) is also called the *substitution formula* for coordinates of a vector. It represents the linear substitution of variables with the matrix  $C'$ , which is the transpose of the matrix consisting of the coefficients  $c'_{ij}$  from (3.34). The matrix  $C'$  is called the *transition matrix* from the basis  $e_1, \dots, e_n$  to the basis  $e'_1, \dots, e'_n$ . In matrix form, formula (3.37) takes the form

$$[\alpha'] = C'[\alpha] \quad \text{or} \quad \alpha' = \alpha C'^*. \quad (3.38)$$

Using formulas (3.36) and (3.38), one easily establishes the connection between  $C$  and  $C'$ .

**Lemma 3.58** *The transition matrices  $C$  and  $C'$  between any two bases of a vector space are nonsingular and are the inverses of each other. That is,  $C' = C^{-1}$ .*

*Proof* Substituting the expression  $[\alpha'] = C'[\alpha]$  into  $[\alpha] = C[\alpha']$ , taking into account the associativity of matrix multiplication, we obtain the equality  $[\alpha] =$

$(CC')[\alpha]$ . This equality holds for all column vectors  $[\alpha]$  of a given length  $n$ , and therefore, the matrix  $CC'$  on the right-hand side is the identity matrix. Indeed, rewriting this equality in the equivalent form  $(CC' - E)[\alpha] = \mathbf{0}$ , it becomes clear that if the matrix  $CC' - E$  contains at least one nonzero element, then there exists a column vector  $[\alpha]$  for which  $(CC' - E)[\alpha] \neq \mathbf{0}$ . Thus we conclude that  $CC' = E$ , from which by definition of the inverse matrix (see Sect. 2.10), it follows that  $C' = C^{-1}$ .  $\square$

We shall now explain how the matrix of a linear transformation depends on the choice of bases. Suppose that in the bases  $e_1, \dots, e_n$  and  $f_1, \dots, f_m$  of the vector spaces  $L$  and  $M$  the transformation  $\mathcal{A} : L \rightarrow M$  has matrix  $A$ , the coordinates of the vector  $x$  are denoted by  $\alpha_i$ , and the coordinates of the vector  $\mathcal{A}(x)$  are denoted by  $\beta_j$ . Similarly, in the bases  $e'_1, \dots, e'_n$  and  $f'_1, \dots, f'_m$  of these vector spaces, the same transformation  $\mathcal{A} : L \rightarrow M$  has matrix  $A'$ , the coordinates of the vector  $x$  are denoted by  $\alpha'_i$ , and the coordinates of the vector  $\mathcal{A}(x)$  are denoted by  $\beta'_j$ .

Let  $C$  be the transition matrix from the basis  $e'_1, \dots, e'_n$  to the basis  $e_1, \dots, e_n$ , which is a nonsingular matrix of order  $n$ , while  $D$  is the transition matrix from the basis  $f'_1, \dots, f'_m$  to the basis  $f_1, \dots, f_m$ , which is a nonsingular matrix of order  $m$  (here  $n$  and  $m$  are the dimensions of the vector spaces  $L$  and  $M$ ). Then by the change-of-coordinates formula (3.38), we obtain

$$[\alpha'] = C^{-1}[\alpha], \quad [\beta'] = D^{-1}[\beta],$$

and formula (3.27) of the linear transformation gives us the equalities

$$[\beta] = A[\alpha], \quad [\beta'] = A'[\alpha'].$$

Let us substitute on the right-hand side of the equality  $[\beta'] = D^{-1}[\beta]$ , the expression  $[\beta] = A[\alpha]$ , and on the left-hand side, the expression  $[\beta'] = A'[\alpha'] = A'C^{-1}[\alpha]$ , as a result of which we obtain the relationship

$$A'C^{-1}[\alpha] = D^{-1}A[\alpha]. \quad (3.39)$$

This line of argument holds for any vector  $x \in L$ , and hence equality (3.39) holds for any column vector  $[\alpha]$  of length  $n$ . Clearly, this is possible if and only if we have the equality

$$A'C^{-1} = D^{-1}A. \quad (3.40)$$

Indeed, both matrices  $A'C^{-1}$  and  $D^{-1}A$  are of type  $(m, n)$ , and if they were not equal, then there would be at least one row (with index  $i$  between 1 and  $n$ ) and one column (with index  $j$  between 1 and  $m$ ) such that the  $ij$ th elements of the matrices  $A'C^{-1}$  and  $D^{-1}A$  did not coincide. But then one could easily identify a column vector  $[\alpha]$  for which the equality (3.39) was not satisfied. For example, set its element  $\alpha_j$  equal to 1, and all the rest to zero.

Let us note that we could have obtained formula (3.40) by considering the transition from one basis to another as a linear transformation of vector spaces given



by multiplication by the transition matrix (see Remark 3.57 above). Indeed, in this case, we obtain the following diagram, in which each arrow indicates multiplication of a column vector by the matrix next to it:

$$\begin{array}{ccc} [\alpha] & \xrightarrow{C^{-1}} & [\alpha'] \\ A \downarrow & & \downarrow A' \\ [\beta] & \xrightarrow{D^{-1}} & [\beta'] \end{array}$$

By the definition of matrix multiplication, from the vector  $[\alpha]$ , we can obtain the vector  $[\beta']$  located in the opposite corner of the diagram in two ways: multiplication by the matrix  $A'C^{-1}$  and multiplication by the matrix  $D^{-1}A$ . Both methods should give the same result (in such case, we say that the diagram is *commutative*, and this is equivalent to equality (3.40)).

We can multiply both sides of (3.40) on the right by the matrix  $C$ , obtaining as a result

$$A' = D^{-1}AC, \quad (3.41)$$

which is called the *formula for a change of matrix of a linear transformation*.

In the case that the dimensions  $n$  and  $m$  of the vector spaces  $L$  and  $M$  coincide, both matrices  $A$  and  $A'$  are square (of order  $n = m$ ), and for such matrices, one has the notion of the determinant. Then by Theorem 2.54, from formula (3.41), there follows the relationship

$$|A'| = |D^{-1}| \cdot |A| \cdot |C| = |D|^{-1} \cdot |A| \cdot |C|. \quad (3.42)$$

Since  $C$  and  $D$  are transition matrices, they are nonsingular, and therefore the determinants  $|A'|$  and  $|A|$  differ from each other through multiplication by the number  $|D|^{-1}|C| \neq 0$ . This indicates that if the matrix of a linear transformation of spaces of the same dimension is nonsingular for some choice of bases, then it will be nonsingular for any other choice of bases for these spaces. Therefore, we may make the following definition.

**Definition 3.59** A linear transformation of spaces of the same dimension is said to be *nonsingular* if its matrix (expressed in terms of some choice of bases of the two spaces) is nonsingular.

There is a special case, which is of greatest importance for a variety of applications to which Chaps. 4 and 5 will be devoted, in which the spaces  $L$  and  $M$  coincide (that is,  $\mathcal{A}$  is a linear transformation of a vector space into itself and so  $n = m$ ), the basis  $e_1, \dots, e_n$  coincides with the basis  $f_1, \dots, f_m$ , and the basis  $e'_1, \dots, e'_n$  coincides with  $f'_1, \dots, f'_m$ . Consequently, in this case,  $D = C$ , and the change-of-matrix formula (3.41) is converted to

$$A' = C^{-1}AC, \quad (3.43)$$

and equation (3.42) assumes the very simple form  $|A'| = |A|$ . This means that although the matrix of a linear transformation of a vector space  $L$  into itself depends on the choice of basis, its determinant does not depend on the choice of basis. This circumstance is frequently expressed by saying that the determinant is *invariant* under a linear transformation of a vector space into itself. In this case, we may give the following definition.

**Definition 3.60** The *determinant* of a linear transformation  $\mathcal{A} : L \rightarrow L$  of a vector space to itself (denoted by  $|\mathcal{A}|$ ) is the determinant of its matrix  $A$ , expressed in terms of any basis of the space  $L$ , that is,  $|\mathcal{A}| = |A|$ .

### 3.5 Isomorphisms of Vector Spaces

In this section we shall investigate the case in which a linear transformation  $\mathcal{A} : L \rightarrow M$  is a *bijection*. We observe first of all that if  $\mathcal{A}$  is a bijective linear transformation from  $L$  to  $M$ , then like any bijective mapping (not necessarily linear), it has an inverse mapping  $\mathcal{A}^{-1} : M \rightarrow L$ . It is clear that  $\mathcal{A}^{-1}$  will also be a linear transformation from  $M$  to  $L$ . Indeed, if for the vector  $y_1 \in M$  there is a unique vector  $x_1 \in L$  such that  $\mathcal{A}(x_1) = y_1$ , and for  $y_2 \in M$  there is an analogous vector  $x_2 \in L$  such that  $\mathcal{A}(x_1 + x_2) = y_1 + y_2$ , then by the definition of inverse mapping, we obtain the first of conditions (3.21) in the definition of a linear transformation:

$$\mathcal{A}^{-1}(y_1 + y_2) = x_1 + x_2 = \mathcal{A}^{-1}(y_1) + \mathcal{A}^{-1}(y_2).$$

Similarly, but even more simply, we can verify the second condition of (3.21), that is, that  $\mathcal{A}^{-1}(\alpha y) = \alpha \mathcal{A}^{-1}(y)$  for an arbitrary vector  $y \in M$  and scalar  $\alpha$ .

**Definition 3.61** Vector spaces  $L$  and  $M$  between which there exists a bijective linear transformation  $\mathcal{A}$  are said to be *isomorphic*, and the transformation  $\mathcal{A}$  itself is called an *isomorphism*. The fact that vector spaces  $L$  and  $M$  are isomorphic is denoted by  $L \simeq M$ . If we wish to specify a concrete transformation  $\mathcal{A} : L \rightarrow M$  that produces the isomorphism, then we write  $\mathcal{A} : L \xrightarrow{\sim} M$ .

The property of being isomorphic defines an equivalence relation on the set of all vector spaces (see the definition on p. xii). To prove this, we need to verify three properties: reflexivity, symmetry, and transitivity. Reflexivity is obvious: we have simply to consider the identity mapping  $\mathcal{E} : L \xrightarrow{\sim} L$ . Symmetry is also obvious: if  $\mathcal{A} : L \xrightarrow{\sim} M$ , then the inverse transformation  $\mathcal{A}^{-1}$  is also an isomorphism, that is,  $\mathcal{A}^{-1} : M \xrightarrow{\sim} L$ . Finally, if  $\mathcal{A} : L \xrightarrow{\sim} M$  and  $\mathcal{B} : M \xrightarrow{\sim} N$ , then, as is easily verified, the transformation  $\mathcal{C} = \mathcal{B}\mathcal{A}$  is also an isomorphism, that is,  $\mathcal{C} : L \xrightarrow{\sim} N$ , which establishes transitivity. Therefore, the set of all vector spaces can be represented as a collection of equivalence classes of vector spaces whose elements are mutually isomorphic.

*Example 3.62* With the choice of basis  $e_1, \dots, e_n$  in a vector space  $L$  over a field  $\mathbb{K}$ , assigning to a vector  $x \in L$  the row consisting of its coordinates in this basis establishes an isomorphism between  $L$  and the row space  $\mathbb{K}^n$ . Similarly, the elements of a row in the form of a column produces an isomorphism between the row space and the column space (with rows and columns containing the same numbers of elements). This explains why we use a single symbol for denoting these spaces.

*Example 3.63* Through the selection of bases  $e_1, \dots, e_n$  and  $f_1, \dots, f_m$  in the spaces  $L$  and  $M$  of dimensions  $n$  and  $m$ , we assign to each linear transformation  $\mathcal{A} : L \rightarrow M$  its matrix  $A$ . We thereby establish an isomorphism between the space  $\mathcal{L}(L, M)$  and the space of rectangular matrices of type  $(m, n)$ .

**Theorem 3.64** *Two finite-dimensional vector spaces  $L$  and  $M$  are isomorphic if and only if  $\dim L = \dim M$ .*

*Proof* The fact that all vector spaces of a given finite dimension are isomorphic follows easily from the fact that every vector space  $L$  of finite dimension  $n$  is isomorphic to the space  $\mathbb{K}^n$  of rows or columns of length  $n$  (Example 3.62). Indeed, let  $L$  and  $M$  be two vector spaces of dimension  $n$ . Then  $L \simeq \mathbb{K}^n$  and  $M \simeq \mathbb{K}^n$ , from which as a result of transitivity and symmetry, we obtain  $L \simeq M$ .

We now prove that isomorphic vector spaces  $L$  and  $M$  have the same dimension. Let us assume that  $\mathcal{A} : L \xrightarrow{\sim} M$  is an isomorphism. Let us denote by  $\mathbf{0} \in L$  and  $\mathbf{0}' \in M$  the null vectors in the spaces  $L$  and  $M$ . Recall, by the property of linear transformations that we proved on p. 102, that  $\mathcal{A}(\mathbf{0}) = \mathbf{0}'$ . Let  $\dim M = m$ , and let us choose in  $M$  some basis  $f_1, \dots, f_m$ . By the definition of isomorphism of a vector space  $L$ , there exist vectors  $e_1, \dots, e_m$  such that  $f_i = \mathcal{A}(e_i)$  for  $i = 1, \dots, m$ . We shall prove that the vectors  $e_1, \dots, e_m$  form a basis of the space  $L$ , whence it will follow that  $\dim L = m$ , completing the proof of the theorem.

First of all, let us show that these vectors are linearly independent. Indeed, if  $e_1, \dots, e_m$  were linearly dependent, then there would exist scalars  $\alpha_1, \dots, \alpha_m$ , not all equal to zero, such that

$$\alpha_1 e_1 + \alpha_2 e_2 + \dots + \alpha_m e_m = \mathbf{0}.$$

But after applying the linear transformation  $\mathcal{A}$  to both parts of this relationship, in view of the equality  $\mathcal{A}(\mathbf{0}) = \mathbf{0}'$ , we would obtain

$$\alpha_1 f_1 + \alpha_2 f_2 + \dots + \alpha_m f_m = \mathbf{0}',$$

from which follows  $\alpha_1 = 0, \dots, \alpha_m = 0$ , since by assumption, the vectors  $f_1, \dots, f_m$  are linearly independent.

Let us now prove that every vector  $x \in L$  is a linear combination of the vectors  $e_1, \dots, e_m$ . Let us set  $\mathcal{A}(x) = y$  and express  $y$  in the form

$$y = \alpha_1 f_1 + \alpha_2 f_2 + \dots + \alpha_m f_m.$$

Applying to both sides of this equality the linear transformation  $\mathcal{A}^{-1}$ , we obtain

$$\mathbf{x} = \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \cdots + \alpha_m \mathbf{e}_m,$$

as required. We have thus shown that the vectors  $\mathbf{e}_1, \dots, \mathbf{e}_m$  form a basis of the vector space  $L$ .  $\square$

*Example 3.65* Suppose we are given a system of  $m$  homogeneous linear equations in  $n$  unknowns  $x_1, \dots, x_n$  and with coefficients in the field  $\mathbb{K}$ . As we saw in Example 3.8 (p. 84), its solution forms a subspace  $L'$  of the space  $\mathbb{K}^n$  of rows of length  $n$ . Since we know that the dimension of the space  $\mathbb{K}^n$  is  $n$ , it follows that  $\dim L' \leq n$ . Let us determine this dimension. To this end, using Theorem 1.15, let us bring our system into echelon form (1.18). Since the equations of the original system are homogeneous, it follows that in (1.18), all the equations will also be homogeneous, that is, all the constant terms  $\bar{b}_i$  are equal to 0. Let  $r$  be the number of principal unknowns, and hence  $(n - r)$  is the number of free unknowns. As shown following the proof of Theorem 1.15, we shall obtain all the solutions of our system by assigning arbitrary values to the free unknowns and then determining the principal unknowns from the first  $r$  equations. That is, if  $(x_1, \dots, x_n)$  is some solution, then comparing to it the row of values of the free unknowns  $(x_{i_1}, \dots, x_{i_{n-r}})$ , we obtain a bijection between the set of solutions of the system and rows of length  $n - r$ . An obvious verification shows that this relationship is an isomorphism of the spaces  $\mathbb{K}^{n-r}$  and  $L'$ . Since  $\dim \mathbb{K}^{n-r} = n - r$ , then by Theorem 3.64, the dimension of the space  $L'$  is also equal to  $n - r$ . Finally, we observe that the number  $r$  is equal to the rank of the matrix of the system (see Sect. 2.8). Therefore, we have obtained the following result: the space of solutions of a homogeneous linear system of equations has dimension  $n - r$ , where  $n$  is the number of unknowns, and  $r$  is the rank of the matrix of the system.

Let  $\mathcal{A} : L \xrightarrow{\sim} M$  be an isomorphism of vector spaces  $L$  and  $M$  of dimension  $n$ , and let  $\mathbf{e}_1, \dots, \mathbf{e}_n$  be a basis of  $L$ . Then the vectors  $\mathcal{A}(\mathbf{e}_1), \dots, \mathcal{A}(\mathbf{e}_n)$  are linearly independent. Indeed, if not, we would have the equality

$$\alpha_1 \mathcal{A}(\mathbf{e}_1) + \cdots + \alpha_n \mathcal{A}(\mathbf{e}_n) = \mathcal{A}(\alpha_1 \mathbf{e}_1 + \cdots + \alpha_n \mathbf{e}_n) = \mathbf{0}',$$

from which by the property  $\mathcal{A}(\mathbf{0}) = \mathbf{0}'$  and that fact that  $\mathcal{A}$  is a bijection, we obtain the relationship  $\alpha_1 \mathbf{e}_1 + \cdots + \alpha_n \mathbf{e}_n = \mathbf{0}$ , contradicting the definition of basis. Hence the vectors  $\mathcal{A}(\mathbf{e}_1), \dots, \mathcal{A}(\mathbf{e}_n)$  form a basis of the vector space  $M$ . It is easy to see that in these bases, the matrix of the transformation  $\mathcal{A}$  is the identity matrix of order  $n$ , and the coordinates of an arbitrary vector  $\mathbf{x} \in L$  in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  coincide with the coordinates of the vector  $\mathcal{A}(\mathbf{x})$  in the basis  $\mathcal{A}(\mathbf{e}_1), \dots, \mathcal{A}(\mathbf{e}_n)$ . Consequently, the transformation  $\mathcal{A}$  is nonsingular.

A similar argument easily establishes the converse fact that an arbitrary nonsingular linear transformation  $\mathcal{A} : L \rightarrow M$  of vector spaces of the same dimension is an isomorphism.

**Remark 3.66** Theorem 3.64 shows that all assertions formulated in terms of concepts entering the definition of a vector space are equivalent for all spaces of a given dimension. In other words, there exists a *single, unique* theory of  $n$ -dimensional vector spaces for a given  $n$ . An example of the opposite situation can be found in Euclidean geometry and the non-Euclidean geometry of Lobachevsky. It is well known that if we accept all the axioms of Euclid except for the “parallel postulate” (so-called *absolute geometry*), then there are two completely different geometries that satisfy these axioms: Euclid’s and Lobachevsky’s. With vector spaces, such a situation does not arise.

The definition of an isomorphism under the linear transformation  $\mathcal{A} : L \rightarrow M$  consists of two parts. The first asserts that for an *arbitrary* vector  $y \in M$ , there exists a vector  $x \in L$  such that  $\mathcal{A}(x) = y$ , that is, the image  $\mathcal{A}(L)$  coincides with the entire space  $M$ . The second condition is that the equality  $\mathcal{A}(x_1) = \mathcal{A}(x_2)$  holds *only* for  $x_1 = x_2$ . Since  $\mathcal{A}$  is a linear transformation, then for the latter condition to be satisfied, it is necessary that the equality  $\mathcal{A}(x) = \mathbf{0}'$  imply  $x = \mathbf{0}$ . This motivates the following definition.

**Definition 3.67** The set of vectors in the space  $L$  such that  $\mathcal{A}(x) = \mathbf{0}'$  is called the *kernel* of the linear transformation  $\mathcal{A}$ .<sup>5</sup> In other words, the kernel is the preimage of the null vector under the mapping  $\mathcal{A}$ .

It is obvious that the kernel of a linear transformation  $\mathcal{A} : L \rightarrow M$  is a subspace of  $L$ , and that its image  $\mathcal{A}(L)$  is a subspace of  $M$ .

Thus to satisfy the second condition in the definition of a bijection, it is necessary that the kernel  $\mathcal{A}$  consist of the null vector alone. But this condition is sufficient as well. Indeed, if for vectors  $x_1 \neq x_2$  the condition  $\mathcal{A}(x_1) = \mathcal{A}(x_2)$  is satisfied, then subtracting one side of the equality from the other and applying the linearity of the transformation  $\mathcal{A}$ , we obtain  $\mathcal{A}(x_1 - x_2) = \mathbf{0}'$ , that is, the vector  $x_1 - x_2$  is in the kernel of  $\mathcal{A}$ . Therefore, the linear transformation  $\mathcal{A} : L \rightarrow M$  is an isomorphism if and only if its image coincides with all of  $M$  and its kernel is equal to  $(\mathbf{0})$ . We shall now show that if  $\mathcal{A}$  is a linear transformation of spaces of *the same finite dimension*, then an isomorphism results if either one or the other of the two conditions is satisfied.

**Theorem 3.68** If  $\mathcal{A} : L \rightarrow M$  is a linear transformation of vector spaces of the same finite dimension and the kernel of  $\mathcal{A}$  is equal to  $(\mathbf{0})$ , then  $\mathcal{A}$  is an isomorphism.

*Proof* Let  $\dim L = \dim M = n$ . Let us consider a particular basis  $e_1, \dots, e_n$  of the vector space  $L$ . The transformation  $\mathcal{A}$  maps each vector  $e_i$  to some vector  $f_i = \mathcal{A}(e_i)$  of the space  $M$ . Then the vectors  $f_1, \dots, f_n$  are linearly independent, that is,

---

<sup>5</sup>*Translator’s note:* Another name for kernel that the reader may encounter is *null space* (since the kernel is the space of all vectors that map to the null vector).

they form a basis of the space  $M$ . Indeed, from the linearity of the transformation  $\mathcal{A}$ , for arbitrary scalars  $\alpha_1, \dots, \alpha_n$ , we have the equality

$$\mathcal{A}(\alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n) = \alpha_1 \mathbf{f}_1 + \dots + \alpha_n \mathbf{f}_n. \quad (3.44)$$

If  $\alpha_1 \mathbf{f}_1 + \dots + \alpha_n \mathbf{f}_n = \mathbf{0}'$  for some collection of scalars  $\alpha_1, \dots, \alpha_n$ , then from the condition that the kernel of  $\mathcal{A}$  is equal to  $\{\mathbf{0}\}$ , we will have  $\alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n = \mathbf{0}$ , from which it follows, by the definition of a basis, that all the scalars  $\alpha_i$  are equal to zero. The relationship (3.44) also shows that the transformation  $\mathcal{A}$  maps each vector  $\mathbf{x} \in L$  with coordinates  $(\alpha_1, \dots, \alpha_n)$  in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  into the vector  $M$  with the same coordinates in the corresponding basis  $\mathbf{f}_1, \dots, \mathbf{f}_n$  (the matrix of the transformation  $\mathcal{A}$  in such bases is the identity matrix of order  $n$ ).

By the definition of an isomorphism, it suffices to prove that for an arbitrary vector  $\mathbf{y} \in M$ , there exists a vector  $\mathbf{x} \in L$  such that  $\mathcal{A}(\mathbf{x}) = \mathbf{y}$ . Since the vectors  $\mathbf{f}_1, \dots, \mathbf{f}_n$  form a basis of the space  $M$ , it follows that  $\mathbf{y}$  can be expressed as a linear combination of these vectors with certain coefficients  $(\alpha_1, \dots, \alpha_n)$ , from which by the linearity of  $\mathcal{A}$  it follows that

$$\mathbf{y} = \alpha_1 \mathbf{f}_1 + \dots + \alpha_n \mathbf{f}_n = \mathcal{A}(\alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n) = \mathcal{A}(\mathbf{x})$$

with vectors  $\mathbf{x} = \alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n$ , which completes the proof of the theorem.  $\square$

**Theorem 3.69** *If  $\mathcal{A} : L \rightarrow M$  is a linear transformation of vector spaces of the same finite dimension and the image of  $\mathcal{A}(L)$  is equal to  $M$ , then  $\mathcal{A}$  is an isomorphism.*

*Proof* Let  $\mathbf{f}_1, \dots, \mathbf{f}_n$  be a basis of the vector space  $M$ . By the condition of the theorem, for each  $\mathbf{f}_i$ , there exists a vector  $\mathbf{e}_i \in L$  such that  $\mathbf{f}_i = \mathcal{A}(\mathbf{e}_i)$ . We shall show that the vectors  $\mathbf{e}_1, \dots, \mathbf{e}_n$  are linearly independent and therefore form a basis of  $L$ . Indeed, if there existed a collection of scalars  $\alpha_1, \dots, \alpha_n$  such that  $\alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n = \mathbf{0}$ , then by  $\mathcal{A}(\mathbf{0}) = \mathbf{0}'$  and the linearity of  $\mathcal{A}$ , we would have the equality

$$\mathcal{A}(\alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n) = \alpha_1 \mathcal{A}(\mathbf{e}_1) + \dots + \alpha_n \mathcal{A}(\mathbf{e}_n) = \alpha_1 \mathbf{f}_1 + \dots + \alpha_n \mathbf{f}_n = \mathbf{0}',$$

from which by the definition of basis it would follow that  $\alpha_i = 0$ . That is, the vectors  $\mathbf{e}_1, \dots, \mathbf{e}_n$  indeed form a basis of the space  $L$ .

It follows from the definition of a basis that an arbitrary vector  $\mathbf{x} \in L$  can be written as  $\mathbf{x} = \alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n$ . From this, we obtain

$$\begin{aligned} \mathcal{A}(\mathbf{x}) &= \mathcal{A}(\alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n) = \alpha_1 \mathcal{A}(\mathbf{e}_1) + \dots + \alpha_n \mathcal{A}(\mathbf{e}_n) \\ &= \alpha_1 \mathbf{f}_1 + \dots + \alpha_n \mathbf{f}_n. \end{aligned}$$

If  $\mathcal{A}(\mathbf{x}) = \mathbf{0}'$ , then we have  $\alpha_1 \mathbf{f}_1 + \dots + \alpha_n \mathbf{f}_n = \mathbf{0}'$ , which is possible only if all the  $\alpha_i$  are equal to 0, since the vectors  $\mathbf{f}_1, \dots, \mathbf{f}_n$  form a basis of the space  $M$ . But then, clearly, the vector  $\mathbf{x} = \alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n$  equals  $\mathbf{0}$ . Therefore, the kernel of the transformation  $\mathcal{A}$  consists solely of the null vector, and by Theorem 3.68,  $\mathcal{A}$  is an isomorphism.  $\square$

It is not difficult to see that the theorems proved just above give us the following result.

**Theorem 3.70** *A linear transformation  $\mathcal{A} : L \rightarrow M$  between vector spaces of the same finite dimension is an isomorphism if and only if it is nonsingular.*

In other words, Theorem 3.70 asserts that for spaces of the same finite dimension, the notion of a nonsingular transformation coincides with that of isomorphism.

With the proof of Theorem 3.68 we have also established one important fact: a nonsingular linear transformation  $\mathcal{A} : L \rightarrow M$  of vector spaces of the same finite dimension maps a basis  $e_1, \dots, e_n$  of the space  $L$  to a basis  $f_1, \dots, f_n$  of the space  $M$ , and every vector  $x \in L$  with coordinates  $(\alpha_1, \dots, \alpha_n)$  in the first basis is mapped to the vector  $\mathcal{A}(x) \in M$  with the same coordinates relative to the second basis. This clearly follows from formula (3.44).

Thus it is possible to define a nonsingular transformation  $\mathcal{A} : L \rightarrow M$  by stating that it maps a particular basis  $e_1, \dots, e_n$  of the space  $L$  into a basis  $f_1, \dots, f_n$  of the space  $M$ , and an arbitrary vector  $x \in L$  with coordinates  $(\alpha_1, \dots, \alpha_n)$  with respect to the basis  $e_1, \dots, e_n$  into the vector of  $M$  with the same coordinates with respect to the basis  $f_1, \dots, f_n$ . Later, we will make use of this method in the case  $L = M$ , when we will be studying certain special subsets  $X \subset L$ , primarily *quadratics*. The basic idea is that subsets  $X$  and  $Y$  are mapped into each other using a certain nonsingular mapping  $\mathcal{A} : L \rightarrow L$  (that is,  $Y = \mathcal{A}(X)$ ) if and only if there exist two bases  $e_1, \dots, e_n$  and  $f_1, \dots, f_n$  of the vector space  $L$  such that the condition of the vector  $x$  belonging to the subset  $X$  in coordinates relative to the basis  $e_1, \dots, e_n$  coincides with the condition of the same vector belonging to  $Y$  in coordinates relative to the basis  $f_1, \dots, f_n$ .

In conclusion, let us return once more to Theorem 1.12, proved in Sect. 1.2, and Corollary 1.13 (Fredholm alternative; see p. 11). This theorem and corollary are now completely obvious, obtained as trivial consequences of a more general result.

Indeed, as we saw in Sect. 2.9, a system of  $n$  linear equations in  $n$  unknowns can be written in matrix form  $A[x] = [b]$ , where  $A$  is a square matrix of order  $n$ ,  $[x]$  is a column vector consisting of the unknowns  $x_1, \dots, x_n$ , and  $[b]$  is a column vector consisting of the constants  $b_1, \dots, b_n$ . Let  $\mathcal{A} : L \rightarrow M$  be a linear transformation between vector spaces of the same dimension  $n$ , having for some bases  $e_1, \dots, e_n$  and  $f_1, \dots, f_n$ , the matrix  $A$ . Let  $b \in M$  be the vector whose coordinates in the basis  $f_1, \dots, f_n$  are equal to  $b_1, \dots, b_n$ . Then we can interpret the linear system  $A[x] = [b]$  as equations

$$\mathcal{A}(x) = b \tag{3.45}$$

with the unknown vector  $x \in L$  whose coordinates in the basis  $e_1, \dots, e_n$  give the solution  $(x_1, \dots, x_n)$  to this system.

We have the following obvious alternative: Either the linear transformation  $\mathcal{A} : L \rightarrow M$  is an isomorphism, or else it is not. By Theorem 3.70, the first case is equivalent to the mapping  $\mathcal{A}$  being nonsingular. Then the kernel of  $\mathcal{A}$  is equal to  $(0)$ , and we have the image  $\mathcal{A}(L) = M$ . Consequently, for an arbitrary vector  $b \in M$ ,

there exists (and indeed, it is unique) a vector  $\mathbf{x} \in L$  such that  $\mathcal{A}(\mathbf{x}) = \mathbf{b}$ , that is, equation (3.45) is solvable. In particular, from this we obtain Theorem 1.12 and its corollary. In the second case, the kernel of  $\mathcal{A}$  contains a nontrivial vector (the associated homogeneous system has a nontrivial solution), and the image  $\mathcal{A}(L)$  is not all of the space  $M$ , that is, there exists a vector  $\mathbf{b} \in M$  such that equation (3.45) is not satisfied (the system  $A[\mathbf{x}] = [\mathbf{b}]$  is inconsistent).

This assertion, that either equation (3.45) has a solution for every right-hand side or the associated homogeneous equation has a nontrivial solution, holds also in the case that  $\mathcal{A}$  is a linear transformation (operator) in an infinite-dimensional space satisfying a certain special condition. Such transformations occur in particular in the theory of integral equations, where this assertion is given the name *Fredholm alternative*.

### 3.6 The Rank of a Linear Transformation

In this section we shall look at linear transformations  $\mathcal{A} : L \rightarrow M$  without making any assumptions about the dimensions  $n$  and  $m$  of the spaces  $L$  and  $M$  except to assume that they are finite. We note that if  $\mathbf{e}_1, \dots, \mathbf{e}_n$  is any basis of the space  $L$ , then the image of  $\mathcal{A}$  is equal to  $\langle \mathcal{A}(\mathbf{e}_1), \dots, \mathcal{A}(\mathbf{e}_n) \rangle$ . If we choose some basis  $\mathbf{f}_1, \dots, \mathbf{f}_m$  of the space  $M$  and write the matrix of the transformation  $\mathcal{A}$  with respect to the chosen bases, then its columns will consist of the coordinates of the vectors  $\mathcal{A}(\mathbf{e}_1), \dots, \mathcal{A}(\mathbf{e}_n)$  in the bases  $\mathbf{f}_1, \dots, \mathbf{f}_m$ , and therefore, the dimension of the image of  $\mathcal{A}$  is equal to the greatest number of linearly independent vectors among these columns, that is, the rank of the matrix of the linear transformation  $\mathcal{A}$ . Thus the rank of the matrix of a linear transformation is independent of the bases in which it is written, and therefore, we may speak of *the rank of a linear transformation*. This allows us to give an equivalent definition of the rank of a linear transformation that does not depend on the choice of coordinates.

**Definition 3.71** The *rank* of a linear transformation  $\mathcal{A} : L \rightarrow M$  is the dimension of the vector space  $\mathcal{A}(L)$ .

The following theorem establishes a connection between the rank of a linear transformation and the dimension of its kernel, and it shows a very simple form into which the matrix of a linear transformation  $\mathcal{A} : L \rightarrow M$  can be brought through a suitable choice of bases of both spaces.

**Theorem 3.72** For any linear transformation  $\mathcal{A} : L \rightarrow M$  of finite-dimensional vector spaces, the dimension of the kernel of  $\mathcal{A}$  is equal to  $\dim L - r$ , where  $r$  is the rank of  $\mathcal{A}$ . In the two spaces, it is possible to choose bases in which the transformation  $\mathcal{A}$  has a matrix in block-diagonal form

$$\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}, \quad (3.46)$$

where  $E_r$  is the identity matrix of order  $r$ .



*Proof* Let us denote the kernel of the transformation  $\mathcal{A}$  by  $L'$ , and its image  $\mathcal{A}(L)$  by  $M'$ . We begin by proving the relationship

$$\dim L' + \dim M' = \dim L. \quad (3.47)$$

By the definition of the rank of a transformation, we have here  $r = \dim M'$ , and thus the equality (3.47) gives precisely the first assertion of the theorem.

Let us consider the mapping  $\mathcal{A}' : L \rightarrow M'$  that assigns to each vector  $x \in L$  the vector  $y = \mathcal{A}(x)$  in  $M'$ , which by assumption is the image of the mapping  $\mathcal{A} : L \rightarrow M$ . It is clear that such a mapping  $\mathcal{A}' : L \rightarrow M'$  is also a linear transformation. In view of Corollary 3.31, we have the decomposition

$$L = L' \oplus L'', \quad (3.48)$$

where  $L''$  is some subspace of  $L$ . We now consider the restriction of the transformation  $\mathcal{A}'$  to the subspace  $L''$  and denote it by  $\mathcal{A}'' : L'' \rightarrow M'$ . It is easily seen that the image of  $\mathcal{A}''$  coincides with the image of  $\mathcal{A}'$ , that is, is equal to  $M'$ . Indeed, since  $M'$  is the image of the original mapping  $\mathcal{A} : L \rightarrow M$ , every vector  $y \in M'$  can be represented in the form  $y = \mathcal{A}(x)$  with some  $x \in L$ . But in view of the decomposition (3.48), we have the equality  $x = u + v$ , where  $u \in L'$  and  $v \in L''$ , and moreover,  $L'$  is the kernel of  $\mathcal{A}$ , that is,  $\mathcal{A}(u) = \mathbf{0}'$ . Consequently,  $\mathcal{A}(x) = \mathcal{A}(u) + \mathcal{A}(v) = \mathcal{A}(v)$ , and this means that the vector  $y = \mathcal{A}(v)$  is the image of the vector  $v \in L''$ .

The kernel of the transformation  $\mathcal{A}'' : L'' \rightarrow M'$  is equal to  $(\mathbf{0})$ . Indeed, by definition, the kernel is equal to  $L' \cap L''$ , and this intersection consists solely of the null vector, since on the right-hand side of the decomposition (3.48) is to be found a direct sum (see Corollary 3.15). As a result, we obtain that the image of the transformation  $\mathcal{A}'' : L'' \rightarrow M'$  is equal to  $M'$ , while its kernel is equal to  $(\mathbf{0})$ , that is, this transformation is an isomorphism. By Theorem 3.64, it follows that  $\dim L'' = \dim M'$ . On the other hand, from the decomposition (3.48) and Theorem 3.41, it follows that  $\dim L' + \dim L'' = \dim L$ . Substituting here  $\dim L''$  by the equal number  $\dim M'$ , we obtain the required equality (3.47).

We shall now prove the assertion of the theorem about bringing the matrix of a linear transformation  $\mathcal{A}$  into the form (3.46). To this end, similar to the decomposition (3.48) of the space  $L$ , we make the decomposition  $M = M' \oplus M''$ , where  $M''$  is some subspace of  $M$ . By the fact proved above that  $\dim L' = n - r$  and in view of (3.48), it follows that  $\dim L'' = r$ . Let us now choose in the subspace  $L''$  some basis  $u_1, \dots, u_r$  and set  $v_i = \mathcal{A}''(u_i)$ , that is, by definition,  $v_i = \mathcal{A}(u_i)$ . As we have seen, the transformation  $\mathcal{A}'' : L'' \rightarrow M'$  is an isomorphism, and therefore, the vectors  $v_1, \dots, v_r$  form a basis of the space  $M'$ , and moreover, in the bases  $u_1, \dots, u_r$  and  $v_1, \dots, v_r$ , the transformation  $\mathcal{A}''$  has the identity  $E_r$  as its matrix.

Let us now choose in the space  $L'$  some basis  $u_{r+1}, \dots, u_n$  and combine it with the basis  $u_1, \dots, u_r$  into the unified basis  $u_1, \dots, u_n$  of the space  $L$ . Similarly, we extend the basis  $v_1, \dots, v_r$  to an arbitrary basis  $v_1, v_2, \dots, v_m$  of the space  $M$ . What will be the matrix of the linear transformation  $\mathcal{A}$  in the constructed bases  $u_1, \dots, u_n$  and  $v_1, \dots, v_m$ ? It is clear that  $\mathcal{A}(u_i) = v_i$  for  $i = 1, \dots, r$  (by construction, for these vectors, the transformation  $\mathcal{A}''$  is the same as  $\mathcal{A}$ ).

On the other hand,  $\mathcal{A}(\mathbf{u}_i) = \mathbf{0}'$  for  $i = r + 1, \dots, n$ , since the vectors  $\mathbf{u}_{r+1}, \dots, \mathbf{u}_n$  are contained in the kernel of  $\mathcal{A}$ . Writing the coordinates of the vectors  $\mathcal{A}(\mathbf{u}_1), \dots, \mathcal{A}(\mathbf{u}_n)$  in the basis  $\mathbf{v}_1, \dots, \mathbf{v}_m$  as the columns of a matrix, we obtain that the matrix of the transformation  $\mathcal{A}$  has the block-diagonal form (3.46).  $\square$

Theorem 3.72 allows us to obtain a simpler and more natural proof of Theorem 2.63 from the previous section.

To this end, we note that every matrix is the matrix of some linear transformation of vector spaces of suitable dimensions, and in particular, a nonsingular square matrix represents an isomorphism of vector spaces of the same dimension. For the matrices  $A$ ,  $B$ , and  $C$  of Theorem 2.63, let us consider the linear transformations  $\mathcal{A} : M \xrightarrow{\sim} M'$ ,  $\mathcal{B} : L' \xrightarrow{\sim} L$ , and  $\mathcal{C} : L \rightarrow M$ , where  $\dim L = \dim L' = n$  and  $\dim M = \dim M' = m$ , having matrices  $A$ ,  $B$ , and  $C$  in some bases.

Let us find the rank of the transformation  $\mathcal{A}\mathcal{C}\mathcal{B} : L' \rightarrow M'$ . From the equalities  $\mathcal{A}(M) = M'$  and  $\mathcal{B}(L') = L$ , it follows that  $\mathcal{A}\mathcal{C}\mathcal{B}(L') = \mathcal{A}(\mathcal{C}(L))$ , whence taking into account the isomorphism  $\mathcal{A}$ , we obtain that  $\dim \mathcal{A}\mathcal{C}\mathcal{B}(L') = \dim \mathcal{C}(L)$ . By definition, the dimension of the image of a linear transformation is equal to its rank, which coincides with the rank of its matrix, written in terms of arbitrary bases, from which it follows that  $\text{rk } ACB = \text{rk } C$ . From this, we finally obtain the required equality  $\text{rk } ACB = \text{rk } C$ .

We would like to emphasize that the matrix of a transformation is reduced to the simple form (3.46) in the case that the spaces  $L$  and  $M$  are different from each other, and it follows that there is no possibility of coordinating their bases, and they are thus chosen independently of each other. We shall see below that in other cases (for example, if  $L = M$ ), there is a more natural way of making this assignment when the bases of the spaces  $L$  and  $M$  are not chosen independently (for example, in the case  $L = M$ , it is simply one and the same basis). Then the question of the simplest form of the matrix of a transformation becomes much more complex.

The statement of Theorem 3.72 on bringing the matrix of a linear transformation into the form (3.46) can be reformulated. As we established in Sect. 3.4 (substitution formula (3.41)), under a change of bases in the spaces  $L$  and  $M$ , the matrix  $A$  of a linear transformation  $\mathcal{A} : L \rightarrow M$  is replaced by the matrix  $A' = D^{-1}AC$ , where  $C$  and  $D$  are the transition matrices for the new bases in the spaces  $L$  and  $M$ . We know that the matrices  $C$  and  $D$  are nonsingular, and conversely, any nonsingular square matrix of the appropriate order can be taken as the transition matrix to a new basis. Therefore, Theorem 3.72 yields the following corollary.

**Corollary 3.73** *For every matrix  $A$  of type  $(m, n)$ , there exist nonsingular square matrices  $C$  and  $D$  of orders  $n$  and  $m$  such that the matrix  $D^{-1}AC$  has the form (3.46).*

### 3.7 Dual Spaces

In this section, we shall examine the notion of a linear transformation  $\mathcal{A} : L \rightarrow M$  in the simplest case of  $\dim M = 1$ . As a result, we shall arrive at a concept very close

to that with which we began our course in Sect. 1.1, but now reformulated more abstractly, in terms of vector spaces. If  $\dim M = 1$ , then after selecting a basis in  $M$  (that is, some nonzero vector  $\mathbf{e}$ ), we can express any vector in this space in the form  $\alpha \mathbf{e}$ , where  $\alpha$  is a scalar (real, complex, or from an arbitrary field  $\mathbb{K}$ , depending on the interpretation that the reader wishes to give to this term). Identifying  $\alpha \mathbf{e}$  with  $\alpha$ , we may consider in place of  $M$  the collection of scalars ( $\mathbb{R}$ ,  $\mathbb{C}$ , or  $\mathbb{K}$ ). In connection with this, we shall in this case denote the vector space  $\mathfrak{L}(L, M)$  introduced in Sect. 3.3 by  $\mathfrak{L}(L, \mathbb{K})$ . It is called the *space of linear functions on  $L$* .

Therefore, a linear function on a space  $L$  is a mapping  $\mathbf{f} : L \rightarrow \mathbb{K}$  that assigns to each vector  $\mathbf{x} \in L$  the number  $\mathbf{f}(\mathbf{x})$  and satisfies the conditions

$$\mathbf{f}(\mathbf{x} + \mathbf{y}) = \mathbf{f}(\mathbf{x}) + \mathbf{f}(\mathbf{y}), \quad \mathbf{f}(\alpha \mathbf{x}) = \alpha \mathbf{f}(\mathbf{x})$$

for all vectors  $\mathbf{x}, \mathbf{y} \in L$  and scalars  $\alpha \in \mathbb{K}$ .

**Example 3.74** If  $L = \mathbb{K}^n$  is the space of rows of length  $n$  with elements in the field  $\mathbb{K}$ , then the notion of linear function introduced above coincides with the concept introduced in Sect. 1.1.

**Example 3.75** Let  $L$  be the space of continuous functions on the interval  $[a, b]$  taking real or complex values. For every function  $\mathbf{x}(t)$  in  $L$ , we set

$$\mathbf{f}_\varphi(\mathbf{x}) = \int_a^b \varphi(t) \mathbf{x}(t) dt, \quad (3.49)$$

where  $\varphi(t)$  is some fixed function in  $L$ . It is clear that  $\mathbf{f}_\varphi(\mathbf{x})$  is a linear function on  $L$ . We observe that in going through all functions  $\varphi(t)$ , we shall obtain by formula (3.49) an infinite number of linear functions on  $L$ , that is, elements of the space  $\mathfrak{L}(L, \mathbb{K})$ , where  $\mathbb{K} = \mathbb{R}$  or  $\mathbb{C}$ . However, it is not possible to obtain all linear functions on  $L$  with the help of formula (3.49). For example, let  $s \in [a, b]$  be some fixed point on this interval. Consider the mapping  $L \rightarrow \mathbb{K}$  that assigns to each function  $\mathbf{x}(t) \in L$  its value at the point  $s$ . It is then clear that such a mapping is a linear function on  $L$ , but it is represented in the form (3.49) for no function  $\varphi(t)$ .

**Definition 3.76** If  $L$  is finite-dimensional, the space  $\mathfrak{L}(L, \mathbb{K})$  is called the *dual to  $L$*  and is denoted by  $L^*$ .

**Remark 3.77** (The infinite-dimensional case) For an infinite-dimensional vector space  $L$  (for example, that considered in Example 3.75 of the space of continuous functions on an interval), the dual space  $L^*$  is defined to be the space not of all linear functions, but only of those satisfying the particular additional condition of *continuity* (in the case of a finite-dimensional space, the requirement of continuity is automatically satisfied).

The study of linear functions on infinite-dimensional vector spaces turns out to be useful in many questions in analysis and mathematical physics. In this direction, the remarkable idea arose to treat arbitrary linear functions as if they had been given

in the form (3.49), where  $\varphi(t)$  is a certain “generalized function” that does not, in general, belong to the initial space  $L$ . This leads to new and interesting results.

For example, if we take as  $L$  the space of functions that are differentiable on the interval  $[a, b]$  and equal to zero at the endpoints, then for a differentiable function  $\varphi(t)$ , the rule of integration by parts can be written in the form  $f_{\varphi'}(x) = -f_{\varphi}(x')$ . But if the derivative  $\varphi'(t)$  does not exist, then it is possible to define a new, “generalized,” function  $\psi(t)$  by  $f_{\psi}(x) = -f_{\varphi}(x')$ . In this case, it is clear that  $\psi(t) = \varphi'(t)$  if the derivative  $\varphi'(t)$  exists and is continuous. Thus it is possible to define derivatives of arbitrary functions (including discontinuous and even generalized functions).

For example, let us suppose that our interval  $[a, b]$  contains in its interior the point 0 and let us calculate the derivative of the function  $h(t)$  that is equal to zero for  $t < 0$  and to 1 for  $t \geq 0$ , and consequently has a discontinuity at the point  $t = 0$ . By definition, for any function  $x(t)$  in  $L$ , we obtain the equality

$$f_{h'}(x) = -f_h(x') = -\int_a^b h(t)x'(t) dt = -\int_0^b x'(t) dt = x(0) - x(b) = x(0),$$

since  $x(b) = 0$ . Consequently, the derivative  $h'(t)$  is a generalized function<sup>6</sup> that assigns to each function  $x(t)$  in  $L$  its value at the point  $t = 0$ .

We now return to exclusive consideration of the finite-dimensional case.

**Theorem 3.78** *If a vector space  $L$  is of finite dimension, then the dual space  $L^*$  has the same dimension.*

*Proof* Let  $e_1, \dots, e_n$  be any basis of the space  $L$ . Let us consider vectors  $f_i \in L^*$ ,  $i = 1, \dots, n$ , where  $f_i$  is defined as a linear function that assigns to a vector

$$x = \alpha_1 e_1 + \alpha_2 e_2 + \dots + \alpha_n e_n \quad (3.50)$$

its  $i$ th coordinate in the basis  $e_1, \dots, e_n$ , that is,

$$f_1(x) = \alpha_1, \quad \dots, \quad f_n(x) = \alpha_n. \quad (3.51)$$

We will thus obtain  $n$  vectors in the dual space. Let us verify that they form a basis of that space.

Let  $f = \beta_1 f_1 + \dots + \beta_n f_n$ . Then applying the function  $f$  to the vector  $x$ , defined by the formula (3.50), we obtain

$$f(x) = \alpha_1 \beta_1 + \alpha_2 \beta_2 + \dots + \alpha_n \beta_n. \quad (3.52)$$

---

<sup>6</sup>Such a generalized function is called a *Dirac delta function* in honor of the English physicist Paul Adrien Maurice Dirac, who was the first to use generalized functions (toward the end of the 1920s) in his work on quantum mechanics.

In particular, assuming  $\mathbf{x} = \mathbf{e}_i$ , we obtain that  $\mathbf{f}(\mathbf{e}_i) = \beta_i$ . Thus the equality  $\mathbf{f} = \mathbf{0}$  (where  $\mathbf{0}$  is the null vector of the space  $L^*$ , that is, a linear function on  $L$  identically equal to zero) means that  $\mathbf{f}(\mathbf{x}) = 0$  for every vector  $\mathbf{x} \in L$ . It is clear that this is the case if and only if  $\beta_1 = 0, \dots, \beta_n = 0$ . By this we have established the linear independence of the functions  $\mathbf{f}_1, \dots, \mathbf{f}_n$ . By equality (3.52), every linear function on  $L$  can be expressed in the form  $\beta_1 \mathbf{f}_1 + \dots + \beta_n \mathbf{f}_n$  with coefficients  $\beta_i = \mathbf{f}(\mathbf{e}_i)$ . This means that the functions  $\mathbf{f}_1, \dots, \mathbf{f}_n$  form a basis of  $L^*$ , from which it follows that  $\dim L = \dim L^* = n$ .  $\square$

The basis  $\mathbf{f}_1, \dots, \mathbf{f}_n$  of the dual space  $L^*$  constructed according to formula (3.51) is called the *dual* to the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the original vector space  $L$ . It is clear that it is defined by the formula

$$\mathbf{f}_i(\mathbf{e}_i) = 1, \quad \mathbf{f}_i(\mathbf{e}_j) = 0 \quad \text{for } j \neq i.$$

We observe that  $L$  and  $L^*$ , like any two finite-dimensional vector spaces of the same dimension, are isomorphic. (For infinite-dimensional vector spaces, this is not in general the case, as in the case examined in Example 3.75 of the space  $L$  of continuous functions on an interval, for which  $L$  and  $L^*$  are not isomorphic.) However, the construction of an isomorphism between them requires the choice of a basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  in  $L$  and a basis  $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n$  in  $L^*$ . Thus between  $L$  and  $L^*$  there does not exist a “natural” isomorphism independent of the choice of basis. If we repeat the process of passage to the dual space twice, we will obtain the space  $(L^*)^*$ , for which it is easy to construct an isomorphism with the original space  $L$  without resorting to the choice of a special basis. The space  $(L^*)^*$  is called the *second dual space* to  $L$  and is denoted by  $L^{**}$ .

Our immediate objective is to define a linear transformation  $\mathcal{A} : L \rightarrow L^{**}$  that is an isomorphism. To do so, we need to define  $\mathcal{A}(\mathbf{x})$  for every vector  $\mathbf{x} \in L$ . The vector  $\mathcal{A}(\mathbf{x})$  must lie in the space  $L^{**}$ , that is, it must be a linear function on the space  $L^*$ . Since  $\mathcal{A}(\mathbf{x})$  is an element of the second dual space  $L^{**}$ , it follows by definition that  $\mathcal{A}(\mathbf{x})$  is a linear transformation that assigns to each element  $\mathbf{f} \in L^*$  (which itself is a linear function on  $L$ ) some number, denoted by  $\mathcal{A}(\mathbf{x})(\mathbf{f})$ . We will define this number by the natural condition

$$\mathcal{A}(\mathbf{x})(\mathbf{f}) = \mathbf{f}(\mathbf{x}) \quad \text{for all } \mathbf{x} \in L, \mathbf{f} \in L^*. \quad (3.53)$$

The transformation  $\mathcal{A}$  is in  $\mathfrak{L}(L, L^{**})$  (its linearity is obvious). To verify that  $\mathcal{A}$  is a bijection, we can use any basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  in  $L$  and the dual basis  $\mathbf{f}_1, \dots, \mathbf{f}_n$  in  $L^*$ . Then, as is easy to verify,  $\mathcal{A}$  is the composition of two isomorphisms: the isomorphism  $L \xrightarrow{\sim} L^*$  constructed in the proof of Theorem 3.78 and the analogous isomorphism  $L^* \xrightarrow{\sim} L^{**}$ , whence it follows that  $\mathcal{A}$  is itself an isomorphism.

The isomorphism  $L \xrightarrow{\sim} L^{**}$  determined by condition (3.53) shows that the vector spaces  $L$  and  $L^*$  play symmetric roles: each of them is the dual of the other. To point out this symmetry more clearly, we shall find it convenient to write the value  $\mathbf{f}(\mathbf{x})$ , whereby  $\mathbf{x} \in L$  and  $\mathbf{f} \in L^*$ , in the form  $(\mathbf{x}, \mathbf{f})$ . The expression  $(\mathbf{x}, \mathbf{f})$  possesses the following easily verified properties:

1.  $(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{f}) = (\mathbf{x}_1, \mathbf{f}) + (\mathbf{x}_2, \mathbf{f})$ ;
2.  $(\mathbf{x}, \mathbf{f}_1 + \mathbf{f}_2) = (\mathbf{x}, \mathbf{f}_1) + (\mathbf{x}, \mathbf{f}_2)$ ;
3.  $(\alpha \mathbf{x}, \mathbf{f}) = \alpha(\mathbf{x}, \mathbf{f})$ ;
4.  $(\mathbf{x}, \alpha \mathbf{f}) = \alpha(\mathbf{x}, \mathbf{f})$ ;
5. if  $(\mathbf{x}, \mathbf{f}) = 0$  for all  $\mathbf{x} \in L$ , then  $\mathbf{f} = \mathbf{0}$ ;
6. if  $(\mathbf{x}, \mathbf{f}) = 0$  for all  $\mathbf{f} \in L^*$ , then  $\mathbf{x} = \mathbf{0}$ .

Conversely, if for two vector spaces  $L$  and  $M$ , the function  $(\mathbf{x}, \mathbf{y})$  is defined, where  $\mathbf{x} \in L$  and  $\mathbf{y} \in M$ , taking numeric values and satisfying conditions (1)–(6), then as is easily verified,  $L \simeq M^*$  and  $M \simeq L^*$ . We shall rely heavily on this fact in Chap. 6 in our study of bilinear forms.

**Definition 3.79** Let  $L'$  be a subspace of the vector space  $L$ . The set of all  $\mathbf{f} \in L^*$  such that  $\mathbf{f}(\mathbf{x}) = 0$  for all  $\mathbf{x} \in L'$  is called the *annihilator* of the subspace  $L'$  and is denoted by  $(L')^a$ .

It follows at once from this definition that  $(L')^a$  is a subspace of  $L^*$ . Let us determine its dimension. Let  $\dim L = n$  and  $\dim L' = r$ . We choose a basis  $\mathbf{e}_1, \dots, \mathbf{e}_r$  of the subspace  $L'$ , extend it to a basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the entire space  $L$ , and consider the dual basis  $\mathbf{f}_1, \dots, \mathbf{f}_n$  of  $L^*$ . From the definition of the dual basis, it follows easily that a linear function  $\mathbf{f} \in L^*$  belongs to  $(L')^a$  if and only if  $\mathbf{f} \in \langle \mathbf{f}_{r+1}, \dots, \mathbf{f}_n \rangle$ . In other words  $(L')^a = \langle \mathbf{f}_{r+1}, \dots, \mathbf{f}_n \rangle$ , and this implies that

$$\dim(L')^a = \dim L - \dim L'. \quad (3.54)$$

If we now consider the natural isomorphism  $L^{**} \xrightarrow{\sim} L$  defined above and with its help identify these spaces, then it is possible to apply the construction given above to the annihilator  $(L')^a$  and examine the obtained subspace  $((L')^a)^a$  in  $L$ . From the definition, it follows that  $L' \subset ((L')^a)^a$ . From the derived relationship (3.54) for dimension, we obtain that  $\dim((L')^a)^a = n - (n - r) = r$ , and by Theorem 3.24, it follows that  $((L')^a)^a = L'$ .

At the same time, we obtain that the subspace  $L'$  consists of all vectors  $\mathbf{x} \in L$  for which

$$\mathbf{f}_{r+1}(\mathbf{x}) = 0, \quad \dots, \quad \mathbf{f}_n(\mathbf{x}) = 0. \quad (3.55)$$

Thus an arbitrary subspace  $L'$  is defined by some system of linear equations (3.55). This fact is well known in the case of lines and planes ( $\dim L = 1, 2$ ) in three-dimensional space from courses in analytic geometry. In the general case, this assertion is the converse of what was proved in Example 3.8 (p. 84).

We have defined the correspondence  $L' \mapsto (L')^a$  between subspaces  $L' \subset L$  and  $(L')^a \subset L^*$ , which in view of the equality  $((L')^a)^a = L'$  is a bijection. We shall denote this correspondence by  $\varepsilon$  and call it *duality*. Let us now point out some simple properties of this correspondence.

If  $L'$  and  $L''$  are two subspaces of  $L$ , then

$$\varepsilon(L' + L'') = \varepsilon(L') \cap \varepsilon(L''). \quad (3.56)$$

In other words, this means that

$$(\mathbf{L}' + \mathbf{L}'')^a = (\mathbf{L}')^a \cap (\mathbf{L}'')^a. \quad (3.57)$$

Indeed, let  $\mathbf{f} \in (\mathbf{L}')^a \cap (\mathbf{L}'')^a$ . By the definition of sum, for every vector  $\mathbf{x} \in \mathbf{L}' + \mathbf{L}''$  we obtain the representation  $\mathbf{x} = \mathbf{x}' + \mathbf{x}''$ , where  $\mathbf{x}' \in \mathbf{L}'$  and  $\mathbf{x}'' \in \mathbf{L}''$ , whence it follows that  $\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}') + \mathbf{f}(\mathbf{x}'') = 0$ , since  $\mathbf{f} \in (\mathbf{L}')^a$  and  $\mathbf{f} \in (\mathbf{L}'')^a$ . Consequently,  $\mathbf{f} \in (\mathbf{L}' + \mathbf{L}'')^a$ , and thus we have proved the inclusion  $(\mathbf{L}')^a \cap (\mathbf{L}'')^a \subset (\mathbf{L}' + \mathbf{L}'')^a$ . Let us now prove the reverse inclusion. Let  $\mathbf{f} \in (\mathbf{L}' + \mathbf{L}'')^a$ , that is,  $\mathbf{f}(\mathbf{x}) = 0$  for all vectors  $\mathbf{x} = \mathbf{x}' + \mathbf{x}''$ , where  $\mathbf{x}' \in \mathbf{L}'$  and  $\mathbf{x}'' \in \mathbf{L}''$ ; in particular, for all vectors in both subspaces  $\mathbf{L}'$  and  $\mathbf{L}''$ , that is, by the definition of the annihilator, we obtain the relationship  $\mathbf{f} \in (\mathbf{L}')^a$  and  $\mathbf{f} \in (\mathbf{L}'')^a$ . Thus  $\mathbf{f} \in (\mathbf{L}')^a \cap (\mathbf{L}'')^a$ , that is,  $(\mathbf{L}' + \mathbf{L}'')^a \subset (\mathbf{L}')^a \cap (\mathbf{L}'')^a$ . From this, by the previous inclusion, we obtain relationship (3.57), and hence the relationship (3.56).

As a result, we may formulate the following almost obvious *duality principle*. Later, we shall prove deeper versions of this principle.

**Proposition 3.80** (Duality principle) *If for all vector spaces of a given finite dimension  $n$  over a given field  $\mathbb{K}$ , a theorem is proven in whose formulation there appear only the notions of subspace, dimension, sum, and intersection, then for all such spaces, a dual theorem holds, obtained from the initial theorem via the following substitution:*

$$\begin{array}{l} \text{dimension } r \\ \text{intersection } \mathbf{L}' \cap \mathbf{L}'' \\ \text{sum } \mathbf{L}' + \mathbf{L}'' \end{array} \quad \left\| \quad \begin{array}{l} \text{dimension } n - r \\ \text{sum } \mathbf{L}' + \mathbf{L}'' \\ \text{intersection } \mathbf{L}' \cap \mathbf{L}'' \end{array} \right.$$

Finally, we shall examine the linear transformation  $\mathcal{A} : \mathbf{L} \rightarrow \mathbf{M}$ . Here, as with all functions, linear functions are written in reverse order to the order of the sets on which they are defined; see p. xv in the Introduction. Using the notation of that section, we define the set  $T = \mathbb{K}$  and restrict the mapping  $\mathfrak{F}(\mathbf{M}, \mathbb{K}) \rightarrow \mathfrak{F}(\mathbf{L}, \mathbb{K})$  constructed there to the subset  $\mathbf{M}^* \subset \mathfrak{F}(\mathbf{M}, \mathbb{K})$ , the space of linear functions on  $\mathbf{M}$ . We observe that the image  $\mathbf{M}^*$  is contained in the space  $\mathbf{L}^* \subset \mathfrak{F}(\mathbf{L}, \mathbb{K})$ , that is, it consists of linear functions on  $\mathbf{L}$ . We shall denote this mapping by  $\mathcal{A}^*$ . According to the definition on page xv, we define a linear transformation  $\mathcal{A}^* : \mathbf{M}^* \rightarrow \mathbf{L}^*$  by determining, for each vector  $\mathbf{g} \in \mathbf{M}^*$ , its value from the equality

$$(\mathcal{A}^*(\mathbf{g}))(x) = \mathbf{g}(\mathcal{A}(x)) \quad \text{for all } x \in \mathbf{L}. \quad (3.58)$$

A trivial verification shows that  $\mathcal{A}^*(\mathbf{g})$  is a linear function on  $\mathbf{L}$ , and  $\mathcal{A}^*$  is a linear transformation of  $\mathbf{M}^*$  to  $\mathbf{L}^*$ . The transformation  $\mathcal{A}^*$  thus constructed is called the *dual transformation* of  $\mathcal{A}$ . Using our earlier notation to write  $\mathbf{f}(\mathbf{x})$  as  $(\mathbf{x}, \mathbf{f})$ , we can write the definition (3.58) in the following form:

$$(\mathcal{A}^*(\mathbf{y}), \mathbf{x}) = (\mathbf{y}, \mathcal{A}(\mathbf{x})) \quad \text{for all } \mathbf{x} \in \mathbf{L} \text{ and } \mathbf{y} \in \mathbf{M}^*.$$

Let us choose in the space  $\mathbf{L}$  some basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , and in  $\mathbf{M}$ , a basis  $\mathbf{f}_1, \dots, \mathbf{f}_m$ , and also dual bases  $\mathbf{e}_1^*, \dots, \mathbf{e}_n^*$  in  $\mathbf{L}^*$  and  $\mathbf{f}_1^*, \dots, \mathbf{f}_m^*$  in  $\mathbf{M}^*$ .

**Theorem 3.81** *The matrix of a transformation  $\mathcal{A} : \mathbf{L} \rightarrow \mathbf{M}$  written in terms of arbitrary bases of the spaces  $\mathbf{L}$  and  $\mathbf{M}$  and the matrix of the dual transformation  $\mathcal{A}^* : \mathbf{M}^* \rightarrow \mathbf{L}^*$  written in the dual bases in the spaces  $\mathbf{M}^*$  and  $\mathbf{L}^*$  are transposes of each other.*

*Proof* Let  $A = (a_{ij})$  be the matrix of the transformation  $\mathcal{A}$  in the bases  $\mathbf{e}_1, \dots, \mathbf{e}_n$  and  $\mathbf{f}_1, \dots, \mathbf{f}_m$ . By formula (3.23), this means that

$$\mathcal{A}(\mathbf{e}_i) = \sum_{j=1}^m a_{ji} \mathbf{f}_j, \quad i = 1, \dots, n. \quad (3.59)$$

By the definition of the dual transformation (formula (3.58)), for every linear function  $\mathbf{f} \in \mathbf{L}^*$ , the following equality holds:

$$(\mathcal{A}^*(\mathbf{f}))(\mathbf{e}_i) = \mathbf{f}(\mathcal{A}(\mathbf{e}_i)), \quad i = 1, \dots, n.$$

If  $\mathbf{e}_1^*, \dots, \mathbf{e}_n^*$  is the basis of  $\mathbf{L}^*$  dual to the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of  $\mathbf{L}$ , and  $\mathbf{f}_1^*, \dots, \mathbf{f}_m^*$  is the basis of  $\mathbf{M}^*$  dual to the basis  $\mathbf{f}_1, \dots, \mathbf{f}_m$  of  $\mathbf{M}$ , then  $\mathcal{A}^*(\mathbf{f}_k^*)$  is a linear function on  $\mathbf{L}$ , as defined in (3.58). In particular, applying  $\mathcal{A}^*(\mathbf{f}_k^*)$  to the vector  $\mathbf{e}_i \in \mathbf{L}$ , taking into account (3.58) and (3.59), we obtain

$$(\mathcal{A}^*(\mathbf{f}_k^*))(\mathbf{e}_i) = \mathbf{f}_k^*(\mathcal{A}(\mathbf{e}_i)) = \left( \mathbf{f}_k^*, \sum_{j=1}^m a_{ji} \mathbf{f}_j \right) = \sum_{j=1}^m a_{ji} (\mathbf{f}_k^*, \mathbf{f}_j),$$

and this number is equal to  $a_{ki}$  by the definition of the dual basis. It is obvious that this linear function on  $\mathbf{L}$  is the function  $\sum_{i=1}^n a_{ki} \mathbf{e}_i^*$ . Thus we obtain that the transformation  $\mathcal{A}^*$  assigns the vector  $\mathbf{f}_k^* \in \mathbf{M}^*$  to the vector

$$\mathcal{A}^*(\mathbf{f}_k^*) = \sum_{i=1}^n a_{ki} \mathbf{e}_i^*, \quad k = 1, \dots, m, \quad (3.60)$$

of the space  $\mathbf{L}^*$ . Comparing formulas (3.59) and (3.60), we conclude that in the given bases, the matrix of the transformation  $\mathcal{A}^*$  is equal to  $A^* = (a_{ji})$ , that is, the transpose of the matrix of the transformation  $\mathcal{A}$ .  $\square$

If we are given two linear transformations of vector spaces,  $\mathcal{A} : \mathbf{L} \rightarrow \mathbf{M}$  and  $\mathcal{B} : \mathbf{M} \rightarrow \mathbf{N}$ , then we can define their composition  $\mathcal{B}\mathcal{A} : \mathbf{L} \rightarrow \mathbf{N}$ , which means that its dual transformation is also defined, and is given by  $(\mathcal{B}\mathcal{A})^* : \mathbf{N}^* \rightarrow \mathbf{L}^*$ . From the condition (3.58), an immediate verification easily leads to the relation

$$(\mathcal{B}\mathcal{A})^* = \mathcal{A}^* \mathcal{B}^*. \quad (3.61)$$

Together with Theorem 3.81, we thus obtain a new proof of equality (2.57), and moreover, now no formulas are used; relationship (2.57) is obtained on the basis of general notions.



### 3.8 Forms and Polynomials in Vectors

A natural generalization of the concept of linear function on a vector space is the notion of *form*. It plays an important role in many branches of mathematics and in mechanics and physics.

In the sequel, we shall assume that the vector space  $L$  on which we want to define a form is defined over an arbitrary field  $\mathbb{K}$ . In the space  $L$ , we choose a basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ . Then every vector  $\mathbf{x} \in L$  is uniquely defined by the choice of coordinates  $(x_1, \dots, x_n)$  in the given basis.

**Definition 3.82** A function  $F : L \rightarrow \mathbb{K}$  is called a *polynomial* on the space  $L$  if  $F(\mathbf{x})$  can be written as a polynomial in the coordinates  $x_1, \dots, x_n$  of the vector  $\mathbf{x}$ , that is,  $F(\mathbf{x})$  is a finite sum of expressions of the form

$$cx_1^{k_1} \cdots x_n^{k_n}, \quad (3.62)$$

where  $k_1, \dots, k_n$  are nonnegative integers and the coefficient  $c$  is in  $\mathbb{K}$ . The expression (3.62) is called a *monomial* in the space  $L$ , while the number  $k = k_1 + \cdots + k_n$  is called its *degree*. The *degree* of  $F(\mathbf{x})$  is the maximum over the degrees of the monomials that enter into it with nonzero coefficients  $c$ .

Let us note that for  $n > 1$ , a polynomial  $F(\mathbf{x})$  of degree  $k$  can have *several* different monomials (3.62) of the same degree entering into it with nonzero coefficients  $c$ .

**Definition 3.83** A polynomial  $F(\mathbf{x})$  on a vector space  $L$  is said to be *homogeneous* of degree  $k$  or a *form* of degree  $k$  (or frequently *k-form*) if every monomial entering into  $F(\mathbf{x})$  with nonzero coefficients is of degree  $k$ .

The definitions we have given require a bit of comment; indeed, we introduced them having chosen a particular basis of the space  $L$ , and now we need to show that everything remains as defined under a change of basis; that is, if the function  $F(\mathbf{x})$  is a polynomial (or form) in the coordinates of the vector  $\mathbf{x}$  in one basis, then it should be a polynomial (or form) of the same degree in the coordinates of the vector  $\mathbf{x}$  in any other basis. Indeed, using the formula for changing the coordinates of a vector, that is, substituting relationships (3.35) into (3.62), it is easily seen that under a change of basis, every monomial (3.62) of degree  $k$  is converted to a sum of monomials of the same degree. Consequently, a change of basis transforms the monomial (3.62) of degree  $k$  into a certain form  $F'(\mathbf{x})$  of degree  $k' \leq k$ . The reason for the inequality here is that monomials entering into this form might cancel, resulting in a leading-degree term that is equal to zero. However, it is easy to see that such cannot occur. For example, using back-substitution, that is, substituting relationship (3.37) into the form  $F'(\mathbf{x})$ , we will clearly again obtain the monomial (3.62). Therefore,  $k \leq k'$ . Thus we have established the equality  $k' = k$ . This establishes everything that we needed to prove.

Forms of degree  $k = 0$  are simply the constant functions, which assign to every vector  $\mathbf{x} \in L$  one and the same number. Forms of degree  $k = 1$  are said to be *linear*,

and these are precisely the linear functions on the space  $L$  that we studied in detail in the previous section.

Forms of degree  $k = 2$  are called *quadratic*; they play an especially important role in courses in linear algebra as well as in many other branches of mathematics and physics. In our course, an entire chapter will be devoted to quadratic forms (Chap. 6).

We observe that we have in fact already encountered forms of arbitrary degree, as shown in the following example.

*Example 3.84* Let  $F(\mathbf{x}_1, \dots, \mathbf{x}_m)$  be a multilinear function on  $m$  rows of length  $n$  (see the definition on p. 51). Since the space  $\mathbb{K}^n$  of rows of length  $n$  is isomorphic to every  $n$ -dimensional vector space, we may view  $F(\mathbf{x}_1, \dots, \mathbf{x}_m)$  as a multilinear function in  $m$  vectors of the space  $L$ . Setting all the vectors  $\mathbf{x}_1, \dots, \mathbf{x}_m$  in  $L$  equal to  $\mathbf{x}$ , then by Theorem 2.29, we obtain on the space  $L$  the form  $\widehat{F}(\mathbf{x}) = F(\mathbf{x}, \dots, \mathbf{x})$  of degree  $m$ .

Let us denote by  $F_k(\mathbf{x})$  the sum of all monomials of degree  $k \geq 0$  appearing in the polynomial  $F(\mathbf{x})$  for a given choice of basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ . Thus  $F_k(\mathbf{x})$  is a form of degree  $k$ , and we obtain the expression

$$F(\mathbf{x}) = F_0 + F_1(\mathbf{x}) + \dots + F_m(\mathbf{x}), \quad (3.63)$$

in which  $F_k(\mathbf{x}) = 0$  if there are no terms of degree  $k$ . For every form  $F_k(\mathbf{x})$  of degree  $k$ , the equation

$$F_k(\lambda \mathbf{x}) = \lambda^k F_k(\mathbf{x}) \quad (3.64)$$

is satisfied for every scalar  $\lambda \in \mathbb{K}$  and every vector  $\mathbf{x} \in L$  (clearly, it suffices to verify (3.64) for a monomial). Substituting in relation (3.63) the vector  $\lambda \mathbf{x}$  in place of  $\mathbf{x}$ , we obtain

$$F(\lambda \mathbf{x}) = F_0 + \lambda F_1(\mathbf{x}) + \dots + \lambda^m F_m(\mathbf{x}). \quad (3.65)$$

From this, it follows easily that the forms  $F_i$  in the representation (3.63) are uniquely determined by the polynomial  $F$ .

It is not difficult to see that the totality of all polynomials on the space  $L$  form a vector space, which we shall denote by  $A$ . This notation is connected with the fact that the totality of all polynomials forms not only a vector space, but a richer and more complex algebraic structure called an *algebra*. This means that in addition to the operations of a vector space, in  $A$  is also defined the operation of the product of every pair of elements satisfying certain conditions; see the definition on p. 370. However, we shall not yet use this fact and will continue to view  $A$  solely as a vector space.

Let us note that the space  $A$  is infinite-dimensional. Indeed, it suffices to consider the infinite sequence of forms  $F_k(\mathbf{x}) = x_i^k$ , where  $k$  runs through the natural numbers, and the form  $F_k(\mathbf{x})$  assigns to a vector  $\mathbf{x}$  with coordinates  $(x_1, \dots, x_n)$  the  $k$ th power of its  $i$ th coordinate (the number  $i$  may be fixed).

The totality of forms of fixed degree  $k$  on a space  $L$  forms a subspace  $A_k \subset A$ . Here  $A_0 = \mathbb{K}$ , and  $A_1$  coincides with the space  $L^*$  of linear functions on  $L$ . The decomposition (3.63) could be interpreted as a decomposition of the space  $A$  as the direct sum of an infinite number of subspaces  $A_k$  ( $k = 0, 1, \dots$ ) if we were to define such a notion. In the field of algebra, the accepted name for this is *graded algebra*.

In the remainder of this section we shall look at two examples that use the concepts just introduced. Here we shall use the rules for differentiating functions of several variables (as applied to polynomials), which is something that might be new to some readers. However, reference to the formulas thus obtained will occur only at isolated places in the course, which can be omitted if desired. We present these arguments only to emphasize the connection with other areas of mathematics.

Let us begin with reasoning that uses a certain coordinate system, that is, a choice of some basis in the space  $L$ . For the polynomial  $F(x_1, \dots, x_n)$ , its partial derivatives are defined by  $\partial F / \partial x_i$ , which are again polynomials. It is easy to see that the mapping that assigns to every polynomial  $F \in A$  the polynomial  $\partial F / \partial x_i$  determines a linear transformation  $A \rightarrow A$ , which we denote by  $\partial / \partial x_i$ . From these transformations we obtain new transformations  $A \rightarrow A$  of the form

$$\mathcal{D} = \sum_{i=1}^n P_i \frac{\partial}{\partial x_i}, \quad (3.66)$$

where the  $P_i$  are arbitrary polynomials. Linear transformations of the form (3.66) are called *first-order differential operators*. In analysis and geometry one considers their analogues, whereby the  $P_i$  are functions of a much more general class and the space  $A$  is correspondingly enlarged. From the simplest properties of differentiation, it follows that the linear operators  $\mathcal{D}$  defined by formula (3.66) exhibit the property

$$\mathcal{D}(FG) = F\mathcal{D}(G) + G\mathcal{D}(F) \quad (3.67)$$

for all  $F \in A$  and  $G \in A$ .

Let us show that the converse also holds: an arbitrary linear transformation  $\mathcal{D} : A \rightarrow A$  satisfying condition (3.67) is a first-order differential operator. To this end, we observe first that from the relation (3.67), it follows that  $\mathcal{D}(1) = 0$ . Indeed, setting in (3.67) the polynomial  $F = 1$ , we obtain the equality  $\mathcal{D}(1G) = 1\mathcal{D}(G) + G\mathcal{D}(1)$ . Canceling the term  $\mathcal{D}(G)$  on the left- and right-hand sides, we see that  $G\mathcal{D}(1) = 0$ , and having selected as  $G$  an arbitrary nonzero polynomial (even if only  $G = 1$ ), we obtain  $\mathcal{D}(1) = 0$ .

Let us now determine a linear transformation  $\mathcal{D}' : A \rightarrow A$  according to the formula

$$\mathcal{D}' = \mathcal{D} - \sum_{i=1}^n P_i \frac{\partial}{\partial x_i}, \quad \text{where } P_i = \mathcal{D}(x_i).$$

It is easily seen that  $\mathcal{D}'(1) = 0$  and  $\mathcal{D}'(x_i) = 0$  for all indices  $i = 1, \dots, n$ . We observe as well that the transformation  $\mathcal{D}'$ , like  $\mathcal{D}$ , satisfies the relationship (3.67),

whence it follows that if  $\mathcal{D}(F) = 0$  and  $\mathcal{D}(G) = 0$ , then also  $\mathcal{D}(FG) = 0$ . Therefore,  $\mathcal{D}'(F) = 0$  if the polynomial  $F$  is the product of any two monomials from the collection  $1, x_1, \dots, x_n$ . It is obvious that into the collection of such polynomials enter all monomials of degree two, and consequently, for them we have  $\mathcal{D}'(F) = 0$ .

Proceeding by induction, we can show that  $\mathcal{D}'(F) = 0$  for all monomials in  $A_k$  for all  $k$ , and therefore, this holds in general for all forms  $F_k \in A_k$ . Finally, we recall that an arbitrary polynomial  $F \in A$  is the sum of a finite number of homogeneous polynomials  $F_k \in A_k$ . Therefore,  $\mathcal{D}'(F) = 0$  for all  $F \in A$ , which means that the transformation  $\mathcal{D}$  has the form (3.66).

The relationship (3.67) gives the definition of a first-order differential operator in a way that does not depend on the coordinate system, that is, on the choice of basis  $e_1, \dots, e_n$  of the space  $L$ .

*Example 3.85* Let us consider the differential operator

$$\tilde{\mathcal{D}} = \sum_{i=1}^n x_i \frac{\partial}{\partial x_i}.$$

It is clear that  $\tilde{\mathcal{D}}(x_i) = x_i$  for all  $i = 1, \dots, n$ , from which it follows that for the restriction to the subspace  $A_1 \subset A$ , the linear transformation  $\tilde{\mathcal{D}} : A_1 \rightarrow A_1$  becomes the identity, that is, equal to  $\mathcal{E}$ . We shall prove that for the restriction to the subspace  $A_k \subset A$ , the transformation  $\tilde{\mathcal{D}} : A_k \rightarrow A_k$  coincides with  $k\mathcal{E}$ . We shall proceed by induction on  $k$ . We have already analyzed the case  $k = 1$ , and the case  $k = 0$  is obvious. Consider now polynomials  $x_i G$ , where  $G \in A_{k-1}$  and  $i = 1, \dots, n$ . Then from (3.67), we have the equality  $\tilde{\mathcal{D}}(x_i G) = x_i \tilde{\mathcal{D}}(G) + G \tilde{\mathcal{D}}(x_i)$ . We have seen that  $\tilde{\mathcal{D}}(x_i) = x_i$ , and by induction, we may assume that  $\tilde{\mathcal{D}}(G) = (k-1)G$ . As a result, we obtain the equality

$$\tilde{\mathcal{D}}(x_i G) = x_i (k-1)G + G x_i = k x_i G.$$

But every polynomial  $F \in A_k$  can be written as the sum of polynomials of the form  $x_i G_i$  with suitable  $G_i \in A_{k-1}$ . Thus for an arbitrary polynomial  $F \in A_k$ , we obtain the relationship  $\tilde{\mathcal{D}}(F) = kF$ . Written in coordinates, this takes the form

$$\sum_{i=1}^n x_i \frac{\partial F}{\partial x_i} = kF, \quad F \in A_k, \quad (3.68)$$

and is called *Euler's identity*.

*Example 3.86* Let  $F(\mathbf{x})$  be an arbitrary polynomial on the vector space  $L$ . For a variable  $t \in \mathbb{R}$  and fixed vector  $\mathbf{x} \in L$ , the function  $F(t\mathbf{x})$ , in view of relationships (3.63) and (3.64), is a polynomial in the variable  $t$ . The expression

$$(d_0 F)(\mathbf{x}) = \left. \frac{d}{dt} F(t\mathbf{x}) \right|_{t=0} \quad (3.69)$$

is called the *differential* of the function  $F(\mathbf{x})$  at the point  $\mathbf{0}$ . Let us point out that on the right-hand side of equality (3.69) can be found the ordinary derivative of  $F(t\mathbf{x})$  as a function of the variable  $t \in \mathbb{R}$  at the point  $t = 0$ . On the left-hand side of the equality (3.69) and in the expression “differential of the function at the point  $\mathbf{0}$ ,” the symbol  $\mathbf{0}$  signifies, as usual, the null vector of the space  $L$ .

Let us now verify that  $(d_0 F)(\mathbf{x})$  is a linear function in  $\mathbf{x}$ . To this end, we use equality (3.65) for the polynomial  $F(t\mathbf{x})$ . From the relationship

$$F(t\mathbf{x}) = F_0 + tF_1(\mathbf{x}) + \cdots + t^m F_m(\mathbf{x}),$$

we obtain immediately that

$$\left. \frac{d}{dt} F(t\mathbf{x}) \right|_{t=0} = F_1(\mathbf{x}),$$

where  $F_1(\mathbf{x})$  is a linear function on  $L$ . Thus in the decomposition (3.63) for the polynomial  $F(\mathbf{x})$ , for the second term,  $F_1(\mathbf{x}) = (d_0 F)(\mathbf{x})$ , and therefore  $d_0 F$  is frequently called the *linear part* of the polynomial  $F$ .

We shall give an expression in coordinates for this important function. Using the rules of differentiation for a function of several variables, we obtain

$$\frac{d}{dt} F(t\mathbf{x}) = \sum_{i=1}^n \frac{\partial F}{\partial x_i}(t\mathbf{x}) \frac{d(t x_i)}{dt} = \sum_{i=1}^n \frac{\partial F}{\partial x_i}(t\mathbf{x}) x_i.$$

Setting  $t = 0$ , we obtain from this formula

$$(d_0 F)(\mathbf{x}) = \sum_{i=1}^n \frac{\partial F}{\partial x_i}(\mathbf{0}) x_i. \quad (3.70)$$

The coordinate representation (3.70) for the differential is quite convenient, but it requires the selection of a basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  in the space  $L$  and the notation  $\mathbf{x} = x_1 \mathbf{e}_1 + \cdots + x_n \mathbf{e}_n$ . The expression (3.69) alone shows that  $(d_0 F)(\mathbf{x})$  does not depend on the choice of basis. In analysis, both expressions (3.69) and (3.70) are defined for functions of a much more general class than polynomials.

We note that for polynomials  $F(x_1, \dots, x_n) = x_i$ , we obtain with the help of formula (3.70) the expression  $(d_0 F)(\mathbf{x}) = x_i$ . This indicates that the functions  $(d_0 x_1), \dots, (d_0 x_n)$  form a basis of  $L^*$  dual to the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of  $L$ .

# Chapter 4

## Linear Transformations of a Vector Space to Itself

### 4.1 Eigenvectors and Invariant Subspaces

In the previous chapter we introduced the notion of a linear transformation of a vector space  $L$  into a vector space  $M$ . In this and the following chapters, we shall consider the important special case in which  $M$  coincides with  $L$ , which in this book will always be assumed to be finite-dimensional. Then a linear transformation  $\mathcal{A} : L \rightarrow L$  will be called a linear transformation of the space  $L$  *to itself*, or simply a linear transformation of the space  $L$ . This case is of great importance, since it is encountered frequently in various fields of mathematics, mechanics, and physics. We now recall some previously introduced facts regarding this case. First of all, as before, we shall understand the term *number* or *scalar* in the broadest possible sense, namely as a real or complex number or indeed as an element of any field  $\mathbb{K}$  (of the reader's choosing).

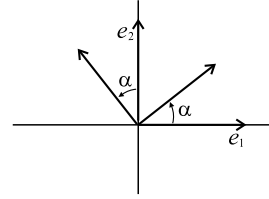
As established in the preceding chapter, to represent a transformation  $\mathcal{A}$  by a matrix, one has to choose a basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the space  $L$  and then to write the coordinates of the vectors  $\mathcal{A}(\mathbf{e}_1), \dots, \mathcal{A}(\mathbf{e}_n)$  in terms of that basis as the columns of a matrix. The result will be a square matrix  $A$  of order  $n$ . If the transformation  $\mathcal{A}$  of the space  $L$  is nonsingular, then the vectors  $\mathcal{A}(\mathbf{e}_1), \dots, \mathcal{A}(\mathbf{e}_n)$  themselves form a basis of the space  $L$ , and we may interpret  $A$  as a transition matrix from the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  to the basis  $\mathcal{A}(\mathbf{e}_1), \dots, \mathcal{A}(\mathbf{e}_n)$ . A nonsingular transformation  $\mathcal{A}$  obviously has an inverse,  $\mathcal{A}^{-1}$ , with matrix  $A^{-1}$ .

*Example 4.1* Let us write down the matrix of the linear transformation  $\mathcal{A}$  that acts by rotating the plane in the counterclockwise direction about the origin through the angle  $\alpha$ . To do so, we first choose a basis consisting of two mutually perpendicular vectors  $\mathbf{e}_1$  and  $\mathbf{e}_2$  of unit length in the plane, where the vector  $\mathbf{e}_2$  is obtained from  $\mathbf{e}_1$  by a counterclockwise rotation through a right angle (see Fig. 4.1).

Then it is easy to see that we obtain the relationship

$$\mathcal{A}(\mathbf{e}_1) = \cos \alpha \mathbf{e}_1 + \sin \alpha \mathbf{e}_2, \quad \mathcal{A}(\mathbf{e}_2) = -\sin \alpha \mathbf{e}_1 + \cos \alpha \mathbf{e}_2,$$

**Fig. 4.1** Rotation through the angle  $\alpha$



and it follows from the definition that the matrix of the transformation  $\mathcal{A}$  in the given basis is equal to

$$A = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}. \quad (4.1)$$

*Example 4.2* Consider the linear transformation  $\mathcal{A}$  of the complex plane that consists in multiplying each number  $z \in \mathbb{C}$  by a given fixed complex number  $p + iq$  (here  $i$  is the imaginary unit).

If we consider the complex plane as a vector space  $L$  over the field  $\mathbb{C}$ , then it is clear that in an arbitrary basis of the space  $L$ , such a transformation  $\mathcal{A}$  has a matrix of order 1, consisting of a unique element, namely the given complex number  $p + iq$ . Thus in this case, we have  $\dim L = 1$ , and we need to choose in  $L$  a basis consisting of an arbitrary nonzero vector in  $L$ , that is, an arbitrary complex number  $z \neq 0$ . Thus we obtain  $\mathcal{A}(z) = (p + iq)z$ .

Now let us consider the complex plane as a vector space  $L$  over the field  $\mathbb{R}$ . In this case,  $\dim L = 2$ , since every complex number  $z = x + iy$  is represented by a pair of real numbers  $x$  and  $y$ . Let us choose in  $L$  the same basis as in Example 4.1. Now we choose the vector  $e_1$  lying on the real axis, and the vector  $e_2$  on the imaginary axis. From the equation

$$(x + iy)(p + iq) = (px - qy) + i(py + qx)$$

it follows that

$$\mathcal{A}(e_1) = pe_1 + qe_2, \quad \mathcal{A}(e_2) = -qe_1 + pe_2,$$

from which it follows by definition that the matrix of the transformation  $\mathcal{A}$  in the given basis takes the form

$$A = \begin{pmatrix} p & -q \\ q & p \end{pmatrix}. \quad (4.2)$$

In the case  $|p + iq| = 1$ , we may put  $p = \cos \alpha$  and  $q = \sin \alpha$  for a certain number  $0 \leq \alpha < 2\pi$  (such an  $\alpha$  is called the *argument* of the complex number  $p + iq$ ). Then the matrix (4.2) coincides with (4.1); that is, multiplication by a complex number with modulus 1 and argument  $\alpha$  is equivalent to the counterclockwise rotation about the origin of the complex plane through the angle  $\alpha$ . We note that every complex number  $p + iq$  can be expressed as the product of a real number  $r$  and a complex

number of modulus 1; that is,  $p + iq = r(p' + iq')$ , where  $|p' + iq'| = 1$  and  $r = |p + iq|$ . From this it is clear that multiplication by  $p + iq$  is the product of two linear transformations of the complex plane: a rotation through the angle  $\alpha$  and a dilation (or contraction) by the factor  $r$ .

In Sect. 3.4, we established that in the transition from a basis  $e_1, \dots, e_n$  of the space  $L$  to some other basis  $e'_1, \dots, e'_n$ , the matrix of the transformation is changed according to the formula

$$A' = C^{-1}AC, \quad (4.3)$$

where  $C$  is the transition matrix from the second basis to the first.

**Definition 4.3** Two square matrices  $A$  and  $A'$  related by (4.3), where  $C$  is any nonsingular matrix, are said to be *similar*.

It is not difficult to see that in the set of square matrices of a given order, the similarity relation thus defined is an equivalence relation (see the definition on p. xii).

It follows from formula (4.3) that in changing bases, the determinant of the transformation matrix does not change, and therefore it is possible to speak not simply about the determinant of the transformation matrix, but about the determinant of the *linear transformation*  $\mathcal{A}$  itself, which will be denoted by  $|\mathcal{A}|$ . A linear transformation  $\mathcal{A} : L \rightarrow L$  is nonsingular if and only if  $|\mathcal{A}| \neq 0$ . If  $L$  is a real space, then this number  $|\mathcal{A}| \neq 0$  is also real and can be either positive or negative.

**Definition 4.4** A nonsingular linear transformation  $\mathcal{A} : L \rightarrow L$  of the real space  $L$  is called *proper* if  $|\mathcal{A}| > 0$ , and *improper* if  $|\mathcal{A}| < 0$ .

One of the basic tasks in the theory of linear transformations, one with which we shall be occupied in the sequel, is to find, given a linear transformation of a vector space into itself, a basis for which the matrix of the transformation takes the simplest possible form. An equivalent formulation of this task is for a given square matrix to find the simplest matrix that is similar to it. Having such a basis (or similar matrix) gives us the possibility of surveying a number of important properties of the initial linear transformation (or matrix). In its most general form, this problem will be solved in Chap. 5, but at present, we shall examine it for a particular type of linear transformation that is most frequently encountered.

**Definition 4.5** A subspace  $L'$  of a vector space  $L$  is called *invariant* with respect to the linear transformation  $\mathcal{A} : L \rightarrow L$  if for every vector  $x \in L'$ , we have  $\mathcal{A}(x) \in L'$ .

It is clear that according to this definition, the zero subspace ( $\mathbf{0}$ ) and the entire space  $L$  are invariant with respect to any linear transformation  $\mathcal{A} : L \rightarrow L$ . Thus whenever we enumerate the invariant subspaces of a space  $L$ , we shall always mean the subspaces  $L' \subset L$  other than ( $\mathbf{0}$ ) and  $L$ .

**Example 4.6** Let  $L$  be the three-dimensional space studied in courses in analytic geometry consisting of vectors originating at a given fixed point  $O$ , and consider the



transformation  $\mathcal{A}$  that reflects each vector with respect to a given plane  $L'$  passing through the point  $O$ . It is then easy to see that  $\mathcal{A}$  has two invariant subspaces: the plane  $L'$  itself and the straight line  $L''$  passing through  $O$  and perpendicular to  $L'$ .

*Example 4.7* Let  $L$  be the same space as in the previous example, and now let the transformation  $\mathcal{A}$  be a rotation through the angle  $\alpha$ ,  $0 < \alpha < \pi$ , about a given axis  $L'$  passing through  $O$ . Then  $\mathcal{A}$  has two invariant subspaces: the line  $L'$  itself and the plane  $L''$  perpendicular to  $L'$  and passing through  $O$ .

*Example 4.8* Let  $L$  be the same as in the previous example, and let  $\mathcal{A}$  be a homothety, that is,  $\mathcal{A}$  acts by multiplying each vector by a fixed number  $\alpha \neq 0$ . Then it is easy to see that every line and every plane passing through  $O$  is an invariant subspace with respect to the transformation  $\mathcal{A}$ . Moreover, it is not difficult to observe that if  $\mathcal{A}$  is a homothety on an arbitrary vector space  $L$ , then every subspace of  $L$  is invariant.

*Example 4.9* Let  $L$  be the plane consisting of all vectors originating at some point  $O$ , and let  $\mathcal{A}$  be the transformation that rotates a vector about  $O$  through the angle  $\alpha$ ,  $0 < \alpha < \pi$ . Then  $\mathcal{A}$  has no invariant subspace.

It is evident that the restriction of a linear transformation  $\mathcal{A}$  to an invariant subspace  $L' \subset L$  is a linear transformation of  $L'$  into itself. We shall denote this transformation by  $\mathcal{A}'$ , that is,  $\mathcal{A}' : L' \rightarrow L'$  and  $\mathcal{A}'(\mathbf{x}) = \mathcal{A}(\mathbf{x})$  for all  $\mathbf{x} \in L'$ .

Let  $\mathbf{e}_1, \dots, \mathbf{e}_m$  be a basis of the subspace  $L'$ . Then since it consists of linearly independent vectors, it is possible to extend it to a basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the entire space  $L$ . Let us examine how the matrix of the linear transformation  $\mathcal{A}$  appears in this basis. The vectors  $\mathcal{A}(\mathbf{e}_1), \dots, \mathcal{A}(\mathbf{e}_m)$  are expressed as a linear combination of  $\mathbf{e}_1, \dots, \mathbf{e}_m$ ; this is equivalent to saying that  $\mathbf{e}_1, \dots, \mathbf{e}_m$  is the basis of a subspace that is invariant with respect to the transformation  $\mathcal{A}$ . We therefore obtain the system of equations

$$\begin{cases} \mathcal{A}(\mathbf{e}_1) = a_{11}\mathbf{e}_1 + a_{21}\mathbf{e}_2 + \cdots + a_{m1}\mathbf{e}_m, \\ \mathcal{A}(\mathbf{e}_2) = a_{12}\mathbf{e}_1 + a_{22}\mathbf{e}_2 + \cdots + a_{m2}\mathbf{e}_m, \\ \vdots \\ \mathcal{A}(\mathbf{e}_m) = a_{1m}\mathbf{e}_1 + a_{2m}\mathbf{e}_2 + \cdots + a_{mm}\mathbf{e}_m. \end{cases}$$

It is clear that the matrix

$$A' = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{pmatrix} \quad (4.4)$$

is the matrix of the linear transformation  $\mathcal{A}' : L' \rightarrow L'$  in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_m$ . In general, we can say nothing about the vectors  $\mathcal{A}(\mathbf{e}_i)$  for  $i > m$  except that they are

linear combinations of vectors from the basis  $e_1, \dots, e_n$  of the entire space  $L$ . However, we shall represent this by separating out terms that are multiples of  $e_1, \dots, e_m$  (we shall write the associated coefficients as  $b_{ij}$ ) and those that are multiples of the vectors  $e_{m+1}, \dots, e_n$  (here we shall write the associated coefficients as  $c_{ij}$ ). As a result we obtain the matrix

$$A = \begin{pmatrix} A' & B' \\ 0 & C' \end{pmatrix}, \quad (4.5)$$

where  $B'$  is a matrix of type  $(m, n - m)$ ,  $C'$  is a square matrix of order  $n - m$ , and  $0$  is a matrix of type  $(n - m, m)$  all of whose elements are equal to zero.

If it turns out to be possible to find an invariant subspace  $L''$  related to the invariant subspace  $L'$  by  $L = L' \oplus L''$ , then by joining the bases of  $L'$  and  $L''$ , we obtain a basis for the space  $L$  in which the matrix of our linear transformation  $\mathcal{A}$  can be written in the form

$$A = \begin{pmatrix} A' & 0 \\ 0 & C' \end{pmatrix},$$

where  $A'$  is the matrix (4.4) and  $C'$  is the matrix of the linear transformation obtained by restricting the transformation  $\mathcal{A}$  to the subspace  $L''$ . Analogously, if

$$L = L_1 \oplus L_2 \oplus \dots \oplus L_k,$$

where all the  $L_i$  are invariant subspaces with respect to the transformation  $\mathcal{A}$ , then the matrix of the transformation  $\mathcal{A}$  can be written in the form

$$A = \begin{pmatrix} A'_1 & 0 & \dots & 0 \\ 0 & A'_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A'_k \end{pmatrix}, \quad (4.6)$$

where  $A'_i$  is the matrix of the linear transformation obtained by restricting  $\mathcal{A}$  to the invariant subspace  $L_i$ . Matrices of the form (4.6) are called *block-diagonal*.

The simplest case is that of an invariant subspace of dimension 1. This subspace has a basis consisting of a single vector  $e \neq 0$ , and its invariance is expressed by the relationship

$$\mathcal{A}(e) = \lambda e \quad (4.7)$$

for some number  $\lambda$ .

**Definition 4.10** If the relationship (4.7) is satisfied for a vector  $e \neq 0$ , then  $e$  is called an *eigenvector*, and the number  $\lambda$  is called an *eigenvalue* of the transformation  $\mathcal{A}$ .

Given an eigenvalue  $\lambda$ , it is easy to verify that the set of all vectors  $e \in L$  satisfying the relationship (4.7), including here also the zero vector, forms an invariant

subspace of  $L$ . It is called the *eigensubspace* for the eigenvalue  $\lambda$  and is denoted by  $L_\lambda$ .

*Example 4.11* In Example 4.6, the eigenvectors of the transformation  $\mathcal{A}$  are, first of all, all the vectors in the plane  $L'$  (in this case the eigenvalue is  $\lambda = 1$ ), and secondly, every vector on the line  $L''$  (the eigenvalue is  $\lambda = -1$ ). In Example 4.7, the eigenvectors are all vectors lying on the line  $L'$ , and to them correspond the eigenvalue  $\lambda = 1$ . In Example 4.8, every vector in the space is an eigenvector with eigenvalue  $\lambda = \alpha$ . Of course all the vectors that we are speaking about are nonzero vectors.

*Example 4.12* Let  $L$  be the space consisting of all infinitely differentiable functions, and let the transformation  $\mathcal{A}$  be differentiation, that is, it maps every function  $x(t)$  in  $L$  to its derivative  $x'(t)$ . Then the eigenvectors of  $\mathcal{A}$  are the functions  $x(t)$ , not identically zero, that are solutions of the differential equation  $x'(t) = \lambda x(t)$ . One easily verifies that such solutions are the functions  $x(t) = ce^{\lambda t}$ , where  $c$  is an arbitrary constant. It follows that to every number  $\lambda$  there corresponds a one-dimensional invariant subspace of the transformation  $\mathcal{A}$  consisting of all vectors  $x(t) = ce^{\lambda t}$ , and for  $c \neq 0$  these are eigenvectors.

There is a convenient method for finding eigenvalues of a transformation  $\mathcal{A}$  and the associated subspaces. We must first choose an arbitrary basis  $e_1, \dots, e_n$  of the space  $L$  and then search for vectors  $e$  that satisfy relation (4.7), in the form of the linear combination

$$e = x_1 e_1 + x_2 e_2 + \dots + x_n e_n. \quad (4.8)$$

Let the matrix of the linear transformation  $\mathcal{A}$  in the basis  $e_1, \dots, e_n$  be  $A = (a_{ij})$ . Then the coordinates of the vector  $\mathcal{A}(e)$  in the same basis can be expressed by the equations

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n, \\ y_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n, \\ \vdots \\ y_n = a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n. \end{cases}$$

Now we can write down relation (4.7) in the form

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = \lambda x_1, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = \lambda x_2, \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = \lambda x_n, \end{cases}$$

or equivalently,

$$\begin{cases} (a_{11} - \lambda)x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0, \\ a_{21}x_1 + (a_{22} - \lambda)x_2 + \cdots + a_{2n}x_n = 0, \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + (a_{nn} - \lambda)x_n = 0. \end{cases} \quad (4.9)$$

For the coordinates  $x_1, x_2, \dots, x_n$  of the vector (4.8), we obtain a system of  $n$  homogeneous linear equations. By Corollary 2.13, this system will have a nonzero solution if and only if the determinant of its matrix is equal to zero. We may write this condition in the form

$$|A - \lambda E| = 0.$$

Using the formula for the expansion of the determinant, we see that the determinant  $|A - tE|$  is a polynomial in  $t$  of degree  $n$ . It is called the *characteristic polynomial* of the transformation  $\mathcal{A}$ . The eigenvalues of  $\mathcal{A}$  are precisely the zeros of this polynomial.

Let us prove that the characteristic polynomial is independent of the basis in which we write down the matrix of the transformation. It is only after we have accomplished this that we shall have the right to speak of the characteristic polynomial of the transformation itself and not merely of its matrix in a particular basis.

Indeed, as we have seen (formula (4.3)), in another basis we obtain the matrix  $A' = C^{-1}AC$ , where  $|C| \neq 0$ . For this matrix, the characteristic polynomial is

$$|A' - tE| = |C^{-1}AC - tE| = |C^{-1}(A - tE)C|.$$

Using the formula for the multiplication of determinants and the formula for the determinant of an inverse matrix, we obtain

$$|C^{-1}(A - tE)C| = |C^{-1}| \cdot |A - tE| \cdot |C| = |A - tE|.$$

If a space has a basis  $e_1, \dots, e_n$  consisting of eigenvectors, then in this basis, we have  $\mathcal{A}(e_i) = \lambda_i e_i$ . From this, it follows that the matrix of a transformation  $\mathcal{A}$  in this basis has the *diagonal form*

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

This is a special case of (4.6) in which the invariant subspaces  $L_i$  are one-dimensional, that is,  $L_i = \langle e_i \rangle$ . Such linear transformations are called *diagonalizable*.

As the following example shows, not all transformations are diagonalizable.

**Example 4.13** Let  $\mathcal{A}$  be a linear transformation of the (real or complex) plane that in some basis  $\mathbf{e}_1, \mathbf{e}_2$  has the matrix

$$A = \begin{pmatrix} a & b \\ 0 & a \end{pmatrix}, \quad b \neq 0.$$

The characteristic polynomial  $|A - tE| = (t - a)^2$  of this transformation has a unique zero  $t = a$ , of multiplicity 2, to which corresponds the one-dimensional eigensubspace  $\langle \mathbf{e}_1 \rangle$ . From this it follows that the transformation  $\mathcal{A}$  is nondiagonalizable.

This can be proved by another method, using the concept of similar matrices. If the transformation  $\mathcal{A}$  were diagonalizable, then there would exist a nonsingular matrix  $C$  of order 2 that would satisfy the relation  $C^{-1}AC = aE$ , or equivalently, the equation  $AC = aC$ . With respect to the unknown elements of the matrix  $C = (c_{ij})$ , the previous equality gives us two equations,  $bc_{21} = 0$  and  $bc_{22} = 0$ , whence by virtue of  $b \neq 0$ , it follows that  $c_{21} = c_{22} = 0$ , and the matrix  $C$  is thus seen to be singular.

We have seen that the number of eigenvalues of a linear transformation is finite, and it cannot exceed the number  $n$  (the dimension of the space  $L$ ), since they are the zeros of the characteristic polynomial, whose degree is  $n$ .

**Theorem 4.14** *The dimension of the eigensubspace  $L_\lambda \subset L$  associated with the eigenvalue  $\lambda$  is at most the multiplicity of the value  $\lambda$  as a zero of the characteristic polynomial.*

*Proof* Suppose the dimension of the eigensubspace  $L_\lambda$  is  $m$ . Let us choose a basis  $\mathbf{e}_1, \dots, \mathbf{e}_m$  of this subspace and extend it to a basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the entire space  $L$ , in which the matrix of the transformation  $\mathcal{A}$  has the form (4.5). Since by the definition of an eigensubspace,  $\mathcal{A}(\mathbf{e}_i) = \lambda \mathbf{e}_i$  for all  $i = 1, \dots, m$ , it follows that in (4.5), the matrix  $A'$  is equal to  $\lambda E_m$ , where  $E_m$  is the identity matrix of order  $m$ . Then

$$A - tE = \begin{pmatrix} A' - tE_m & B' \\ 0 & C' - tE_{n-m} \end{pmatrix} = \begin{pmatrix} (\lambda - t)E_m & B' \\ 0 & C' - tE_{n-m} \end{pmatrix},$$

where  $E_{n-m}$  is the identity matrix of order  $n - m$ . Therefore,

$$|A - tE| = (\lambda - t)^m |C' - tE_{n-m}|.$$

On the other hand, if  $L = L_\lambda \oplus L''$ , then  $L_\lambda \cap L'' = (\mathbf{0})$ , which means that the restriction of the transformation  $\mathcal{A}$  to  $L''$  has no eigenvectors with eigenvalue  $\lambda$ . This means that  $|C' - \lambda E_{n-m}| \neq 0$ , that is, the number  $\lambda$  is not a zero of the polynomial  $|C' - tE_{n-m}|$ , which is what we had to show.  $\square$

In the previous chapter we were introduced to the operations of addition and multiplication (composition) of linear transformations, which are clearly defined

for the special case of a transformation of a space  $L$  into itself. Therefore, for any integer  $n > 0$  we may define the  $n$ th power of a linear transformation. By definition,  $\mathcal{A}^n$  for  $n > 0$  is the result of multiplying  $\mathcal{A}$  by itself  $n$  times, and for  $n = 0$ ,  $\mathcal{A}^0$  is the identity transformation  $\mathcal{E}$ . This enables us to introduce the concept of a *polynomial in a linear transformation*, which will play an important role in what follows.

Let  $\mathcal{A}$  be a linear transformation of the vector space  $L$  (real, complex, or over an arbitrary field  $\mathbb{K}$ ) and define

$$f(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_k x^k,$$

a polynomial with scalar coefficients (respectively real, complex, or from the field  $\mathbb{K}$ ).

**Definition 4.15** A polynomial  $f$  in the linear transformation  $\mathcal{A}$  is a linear mapping

$$f(\mathcal{A}) = \alpha_0 \mathcal{E} + \alpha_1 \mathcal{A} + \cdots + \alpha_k \mathcal{A}^k, \quad (4.10)$$

where  $\mathcal{E}$  is the identity linear transformation.

We observe that this definition does not make use of coordinates, that is, the choice of a specific basis in the space  $L$ . If such a basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  is chosen, then to the linear transformation  $\mathcal{A}$  there corresponds a unique square matrix  $A$ . In Sect. 2.9 we introduced the notion of a polynomial in a square matrix, which allows us to give another definition:  $f(\mathcal{A})$  is the linear transformation with matrix

$$f(A) = \alpha_0 E + \alpha_1 A + \cdots + \alpha_k A^k \quad (4.11)$$

in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ .

It is not difficult to be convinced of the equivalence of these definitions if we recall that the actions of linear transformations are expressed through the actions of their matrices (see Sect. 3.3). It is thus necessary to show that in a change of basis from  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , the matrix  $f(A)$  also changes according to formula (4.3) with transition matrix  $C$  the same as for matrix  $A$ . Indeed, let us consider a change of coordinates (that is, switching to another basis of the space  $L$ ) with matrix  $C$ . Then in the new basis, the matrix of the transformation  $\mathcal{A}$  is given by  $A' = C^{-1}AC$ . By the associativity of matrix multiplication, we also obtain a relationship  $A'^n = C^{-1}A^n C$  for every integer  $n \geq 0$ . If we substitute  $A'$  for  $A$  in formula (4.11), then considering what we have said, we obtain

$$\begin{aligned} f(A') &= \alpha_0 E + \alpha_1 A' + \cdots + \alpha_k A'^k \\ &= C^{-1}(\alpha_0 E + \alpha_1 A + \cdots + \alpha_k A^k)C = C^{-1}f(A)C, \end{aligned}$$

which proves our assertion.

It should be clear that the statements that we proved in Sect. 2.9 for polynomials in a matrix (p. 69) also apply to polynomials in a linear transformation.

**Lemma 4.16** *If  $f(x) + g(x) = u(x)$  and  $f(x)g(x) = v(x)$ , then for an arbitrary linear transformation  $\mathcal{A}$ , we have*

$$f(\mathcal{A}) + g(\mathcal{A}) = u(\mathcal{A}), \quad (4.12)$$

$$f(\mathcal{A})g(\mathcal{A}) = v(\mathcal{A}). \quad (4.13)$$

**Corollary 4.17** *Polynomials  $f(\mathcal{A})$  and  $g(\mathcal{A})$  in the same linear transformation  $\mathcal{A}$  commute:  $f(\mathcal{A})g(\mathcal{A}) = g(\mathcal{A})f(\mathcal{A})$ .*

## 4.2 Complex and Real Vector Spaces

We shall now investigate in greater detail the concepts introduced in the previous section applied to transformations of complex and real vector spaces (that is, we shall assume that the field  $\mathbb{K}$  is respectively  $\mathbb{C}$  or  $\mathbb{R}$ ). Our fundamental result applies specifically to complex spaces.

**Theorem 4.18** *Every linear transformation of a complex vector space has an eigenvector.*

This follows immediately from the fact that the characteristic polynomial of a linear transformation, and in general an arbitrary polynomial of positive degree, has a complex root. Nevertheless, as Example 4.13 of the previous section shows, even in a complex space, not every linear transformation is diagonalizable.

Let us consider the question of diagonalizability in greater detail, always assuming that we are working with complex spaces. We shall prove the diagonalizability of a commonly occurring type of transformation. To this end, we require the following lemma.

**Lemma 4.19** *Eigenvectors associated with distinct eigenvalues are linearly independent.*

*Proof* Suppose the eigenvectors  $\mathbf{e}_1, \dots, \mathbf{e}_m$  are associated with distinct eigenvalues  $\lambda_1, \dots, \lambda_m$ ,

$$\mathcal{A}(\mathbf{e}_i) = \lambda_i \mathbf{e}_i, \quad i = 1, \dots, m.$$

We shall prove the lemma by induction on the number  $m$  of vectors. For the case  $m = 1$ , the result follows from the definition of an eigenvector, namely that  $\mathbf{e}_1 \neq \mathbf{0}$ .

Let us assume that there exists a linear dependence

$$\alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \dots + \alpha_m \mathbf{e}_m = \mathbf{0}. \quad (4.14)$$

Applying the transformation  $\mathcal{A}$  to both sides of the equation, we obtain

$$\lambda_1 \alpha_1 \mathbf{e}_1 + \lambda_2 \alpha_2 \mathbf{e}_2 + \dots + \lambda_m \alpha_m \mathbf{e}_m = \mathbf{0}. \quad (4.15)$$

Subtracting (4.14) multiplied by  $\lambda_m$  from (4.15), we obtain

$$\alpha_1(\lambda_1 - \lambda_m)\mathbf{e}_1 + \alpha_2(\lambda_2 - \lambda_m)\mathbf{e}_2 + \cdots + \alpha_{m-1}(\lambda_{m-1} - \lambda_m)\mathbf{e}_{m-1} = \mathbf{0}.$$

By our induction hypothesis, we may consider that the lemma has been proved for  $m - 1$  vectors  $\mathbf{e}_1, \dots, \mathbf{e}_{m-1}$ . Thus we obtain that  $\alpha_1(\lambda_1 - \lambda_m) = 0, \dots, \alpha_{m-1}(\lambda_{m-1} - \lambda_m) = 0$ , and since by the condition in the lemma,  $\lambda_1 \neq \lambda_m, \dots, \lambda_{m-1} \neq \lambda_m$ , it follows that  $\alpha_1 = \cdots = \alpha_{m-1} = 0$ . Substituting this into (4.14), we arrive at the relationship  $\alpha_m \mathbf{e}_m = \mathbf{0}$ , that is (by the definition of an eigenvector),  $\alpha_m = 0$ . Therefore, in (4.14), all the  $\alpha_i$  are equal to zero, which demonstrates the linear independence of  $\mathbf{e}_1, \dots, \mathbf{e}_m$ .  $\square$

By Lemma 4.19, we have the following result.

**Theorem 4.20** *A linear transformation on a complex vector space is diagonalizable if its characteristic polynomial has no multiple roots.*

As is well known, in this case, the characteristic polynomial has  $n$  distinct roots (we recall once again that we are speaking about polynomials over the field of complex numbers).

*Proof of Theorem 4.20* Let  $\lambda_1, \dots, \lambda_n$  be the distinct roots of the characteristic polynomial of the transformation  $\mathcal{A}$  and let  $\mathbf{e}_1, \dots, \mathbf{e}_n$  be the corresponding eigenvectors. It suffices to show that these vectors form a basis of the entire space. Since their number is equal to the dimension of the space, this is equivalent to showing their linear independence, which follows from Lemma 4.19.  $\square$

If  $A$  is the matrix of the transformation  $\mathcal{A}$  in some basis, then the condition of Theorem 4.20 is satisfied if and only if the so-called *discriminant* of the characteristic polynomial is nonzero.<sup>1</sup> For example, if the order of a matrix  $A$  is 2, and

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

then

$$|A - tE| = \begin{vmatrix} a-t & b \\ c & d-t \end{vmatrix} = (a-t)(d-t) - bc = t^2 - (a+d)t + ad - bc.$$

The condition that this quadratic trinomial have two distinct roots is that  $(a+d)^2 - 4(ad - bc) \neq 0$ . This can be rewritten in the form

$$(a-d)^2 + 4bc \neq 0. \quad (4.16)$$

---

<sup>1</sup>For the general notion of the discriminant of a polynomial, see, for instance, *Polynomials*, by Victor V. Prasolov, Springer 2004.



Similarly, for complex vector spaces of arbitrary dimension, linear transformations not satisfying the conditions of Theorem 4.20 have a matrix that regardless of the basis, has elements that satisfy a special algebraic relationship. In this sense, only exceptional transformations do not meet the conditions of Theorem 4.20.

Analogous considerations give necessary and sufficient conditions for a linear transformation to be diagonalizable.

**Theorem 4.21** *A linear transformation of a complex vector space is diagonalizable if and only if for each of its eigenvalues  $\lambda$ , the dimension of the corresponding eigenspace  $L_\lambda$  is equal to the multiplicity of  $\lambda$  as a root of the characteristic polynomial.*

In other words, the bound on the dimension of the subspace  $L_\lambda$  obtained in Theorem 4.14 is attained.

*Proof of Theorem 4.21* Let the transformation  $\mathcal{A}$  be diagonalizable, that is, in some basis  $e_1, \dots, e_n$  it has the matrix

$$A = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

It is possible to arrange the eigenvalues  $\lambda_1, \dots, \lambda_n$  so that those that are equal are next to each other, so that altogether, they have the form

$$\underbrace{\lambda_1, \dots, \lambda_1}_{m_1 \text{ times}}, \underbrace{\lambda_2, \dots, \lambda_2}_{m_2 \text{ times}}, \dots, \underbrace{\lambda_k, \dots, \lambda_k}_{m_k \text{ times}},$$

where all the numbers  $\lambda_1, \dots, \lambda_k$  are distinct. In other words, we can write the matrix  $A$  in the block-diagonal form

$$A = \begin{pmatrix} \lambda_1 E_{m_1} & 0 & \cdots & 0 \\ 0 & \lambda_2 E_{m_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k E_{m_k} \end{pmatrix}, \quad (4.17)$$

where  $E_{m_i}$  is the identity matrix of order  $m_i$ . Then

$$|A - tE| = (\lambda_1 - t)^{m_1} (\lambda_2 - t)^{m_2} \cdots (\lambda_k - t)^{m_k},$$

that is, the number  $\lambda_i$  is a root of multiplicity  $m_i$  of the characteristic equation. On the other hand, the equality  $\mathcal{A}(x) = \lambda_i x$  for vectors  $x = \alpha_1 e_1 + \cdots + \alpha_n e_n$  gives the relationship  $\lambda_s \alpha_j = \lambda_i \alpha_j$  for all  $j = 1, \dots, n$  and  $s = 1, \dots, k$ , that is, either  $\alpha_j = 0$  or  $\lambda_s = \lambda_i$ . In other words, the vector  $x$  is a linear combination only

of those eigenvectors  $\mathbf{e}_j$  that correspond to the eigenvalue  $\lambda_i$ . This means that the subspace  $L_{\lambda_i}$  consists of all linear combinations of such vectors, and consequently,  $\dim L_{\lambda_i} = m_i$ .

Conversely, for distinct eigenvalues  $\lambda_1, \dots, \lambda_k$ , let the dimension of the eigensubspace  $L_{\lambda_i}$  be equal to the multiplicity  $m_i$  of the number  $\lambda_i$  as a root of the characteristic polynomial. Then from known properties of polynomials, it follows that  $m_1 + \dots + m_k = n$ , which means that

$$\dim L_{\lambda_1} + \dots + \dim L_{\lambda_k} = \dim L. \quad (4.18)$$

We shall show that the sum  $L_{\lambda_1} + \dots + L_{\lambda_k}$  is a direct sum of its eigensubspaces  $L_{\lambda_i}$ . To do so, it suffices to show that for all vectors  $\mathbf{x}_1 \in L_{\lambda_1}, \dots, \mathbf{x}_k \in L_{\lambda_k}$ , the equality  $\mathbf{x}_1 + \dots + \mathbf{x}_k = \mathbf{0}$  is possible only in the case that  $\mathbf{x}_1 = \dots = \mathbf{x}_k = \mathbf{0}$ . But since  $\mathbf{x}_1, \dots, \mathbf{x}_k$  are eigenvectors of the transformation  $\mathcal{A}$  corresponding to distinct eigenvalues  $\lambda_1, \dots, \lambda_k$ , the required assertion follows by Lemma 4.19. Therefore, by equality (4.18), we have the decomposition

$$L = L_{\lambda_1} \oplus \dots \oplus L_{\lambda_k}.$$

Having chosen from each eigensubspace  $L_{\lambda_i}$ ,  $i = 1, \dots, k$ , a basis (consisting of  $m_i$  vectors), and having ordered them in such a way that the vectors entering into a particular subspace  $L_{\lambda_i}$  are adjacent, we obtain a basis of the space  $L$  in which the matrix  $A$  of the transformation  $\mathcal{A}$  has the form (4.17). This means that the transformation  $\mathcal{A}$  is diagonalizable.  $\square$

The case of real vector spaces is more frequently encountered in applications. Their study proceeds in almost the same way as with complex vector spaces, except that the results are somewhat more complicated. We shall introduce here a proof of the real analogue of Theorem 4.18.

**Theorem 4.22** *Every linear transformation of a real vector space of dimension  $n > 2$  has either a one-dimensional or two-dimensional invariant subspace.*

*Proof* Let  $\mathcal{A}$  be a linear transformation of a real vector space  $L$  of dimension  $n > 2$ , and let  $\mathbf{x} \in L$  be some nonnull vector. Since the collection  $\mathbf{x}, \mathcal{A}(\mathbf{x}), \mathcal{A}^2(\mathbf{x}), \dots, \mathcal{A}^n(\mathbf{x})$  consists of  $n + 1 > \dim L$  vectors, then by the definition of the dimension of a vector space, these vectors must be linearly dependent. This means that there exist real numbers  $\alpha_0, \alpha_1, \dots, \alpha_n$ , not all zero, such that

$$\alpha_0 \mathbf{x} + \alpha_1 \mathcal{A}(\mathbf{x}) + \alpha_2 \mathcal{A}^2(\mathbf{x}) + \dots + \alpha_n \mathcal{A}^n(\mathbf{x}) = \mathbf{0}. \quad (4.19)$$

Consider the polynomial  $P(t) = \alpha_0 + \alpha_1 t + \dots + \alpha_n t^n$  and substitute for the variable  $t$ , the transformation  $\mathcal{A}$ , as was done in Sect. 4.1 (formula (4.10)). Then the equality (4.19) can be written in the form

$$P(\mathcal{A})(\mathbf{x}) = \mathbf{0}. \quad (4.20)$$

A polynomial  $P(t)$  satisfying equality (4.20) is called an *annihilator polynomial* of the vector  $\mathbf{x}$  (where it is implied that it is relative to the given transformation  $\mathcal{A}$ ).

Let us assume that the annihilator polynomial  $P(t)$  of some vector  $\mathbf{x} \neq \mathbf{0}$  is the product of two polynomials of lower degree:  $P(t) = Q_1(t)Q_2(t)$ . Then by definition (4.20) and formula (4.13) from the previous section, we have  $Q_1(\mathcal{A})Q_2(\mathcal{A})(\mathbf{x}) = \mathbf{0}$ . Then either  $Q_2(\mathcal{A})(\mathbf{x}) = \mathbf{0}$ , and hence the vector  $\mathbf{x}$  is annihilated by an annihilator polynomial  $Q_2(t)$  of lower degree, or else  $Q_2(\mathcal{A})(\mathbf{x}) \neq \mathbf{0}$ . If we assume  $\mathbf{y} = Q_2(\mathcal{A})(\mathbf{x})$ , we obtain the equality  $Q_1(\mathcal{A})(\mathbf{y}) = \mathbf{0}$ , which means that the non-null vector  $\mathbf{y}$  is annihilated by the annihilator polynomial  $Q_1(t)$  of lower degree. As is well known, an arbitrary polynomial with real coefficients is a product of polynomials of first and second degree. Applying to  $P(t)$  as many times as necessary the process described above, we finally arrive at a polynomial  $Q(t)$  of first or second degree and a nonnull vector  $\mathbf{z}$  such that  $Q(\mathcal{A})(\mathbf{z}) = \mathbf{0}$ . This is the real analogue of Theorem 4.18.

Factoring out the coefficient of the high-order term of  $Q(t)$ , we may assume that this coefficient is equal to 1. If the degree of  $Q(t)$  is equal to 1, then  $Q(t) = t - \lambda$  for some  $\lambda$ , and the equality  $Q(\mathcal{A})(\mathbf{z}) = \mathbf{0}$  yields  $(\mathcal{A} - \lambda\mathcal{E})(\mathbf{z}) = \mathbf{0}$ . This means that  $\lambda$  is an eigenvalue of  $\mathcal{A}$ , which is an eigenvector of the transformation  $\mathcal{A}$ , and therefore,  $\langle \mathbf{z} \rangle$  is a one-dimensional invariant subspace of the transformation  $\mathcal{A}$ .

If the degree of  $Q(t)$  is equal to 2, then  $Q(t) = t^2 + pt + q$  and  $(\mathcal{A}^2 + p\mathcal{A} + q\mathcal{E})(\mathbf{z}) = \mathbf{0}$ . In this case, the subspace  $L' = \langle \mathbf{z}, \mathcal{A}(\mathbf{z}) \rangle$  is two-dimensional and is invariant with respect to  $\mathcal{A}$ . Indeed, the vectors  $\mathbf{z}$  and  $\mathcal{A}(\mathbf{z})$  are linearly independent, since otherwise, we would have the case of an eigenvector  $\mathbf{z}$  considered above. This means that  $\dim L' = 2$ . We shall show that  $L'$  is an invariant subspace of the transformation  $\mathcal{A}$ . Let  $\mathbf{x} = \alpha\mathbf{z} + \beta\mathcal{A}(\mathbf{z})$ . To show that  $\mathcal{A}(\mathbf{x}) \in L'$ , it suffices to verify that vectors  $\mathcal{A}(\mathbf{z})$  and  $\mathcal{A}(\mathcal{A}(\mathbf{z}))$  belong to  $L'$ . This holds for the former by the definition of  $L'$ . It holds for the latter by the fact that  $\mathcal{A}(\mathcal{A}(\mathbf{z})) = \mathcal{A}^2(\mathbf{z})$  and by the condition of the theorem,  $\mathcal{A}^2(\mathbf{z}) + p\mathcal{A}(\mathbf{z}) + q\mathbf{z} = \mathbf{0}$ , that is,  $\mathcal{A}^2(\mathbf{z}) = -q\mathbf{z} - p\mathcal{A}(\mathbf{z})$ .  $\square$

Let us discuss the concept of the annihilator polynomial that we encountered in the proof of Theorem 4.22. An annihilator polynomial of a vector  $\mathbf{x} \neq \mathbf{0}$  having minimal degree is called a *minimal polynomial* of the vector  $\mathbf{x}$ .

**Theorem 4.23** *Every annihilator polynomial is divisible by a minimal polynomial.*

*Proof* Let  $P(t)$  be an annihilator polynomial of the vector  $\mathbf{x} \neq \mathbf{0}$ , and  $Q(t)$  a minimal polynomial. Let us suppose that  $P$  is not divisible by  $Q$ . We divide  $P$  by  $Q$  with remainder. This gives the equality  $P = UQ + R$ , where  $U$  and  $R$  are polynomials in  $t$ , and moreover,  $R$  is not identically zero, and the degree of  $R$  is less than that of  $Q$ . If we substitute into this equality the transformation  $\mathcal{A}$  for the variable  $t$ , then by formulas (4.12) and (4.13), we obtain that

$$P(\mathcal{A})(\mathbf{x}) = U(\mathcal{A})Q(\mathcal{A})(\mathbf{x}) + R(\mathcal{A})(\mathbf{x}), \quad (4.21)$$

and since  $P$  and  $Q$  are annihilator polynomials of the vector  $\mathbf{x}$ , it follows that  $R(\mathcal{A})(\mathbf{x}) = \mathbf{0}$ . Since the degree of  $R$  is less than that of  $Q$ , this contradicts the minimality of the polynomial  $Q$ .  $\square$

**Corollary 4.24** *The minimal polynomial of a vector  $\mathbf{x} \neq \mathbf{0}$  is uniquely defined up to a constant factor.*

Let us note that for the annihilator polynomial, Theorem 4.23 and its converse hold: any multiple of any annihilator polynomial is also an annihilator polynomial (of course, of the same vector  $\mathbf{x}$ ). This follows from the fact that in this case, in equality (4.21), we have  $R = 0$ . From this follows the assertion that there exists a single polynomial that is an annihilator for all vectors of the space  $L$ . Indeed, let  $\mathbf{e}_1, \dots, \mathbf{e}_n$  be some basis of the space  $L$ , and let  $P_1, \dots, P_n$  be annihilator polynomials for these vectors. Let us denote by  $Q$  the least common multiple of these polynomials. Then from what we have said above, it follows that  $Q$  is an annihilator polynomial for each of the vectors  $\mathbf{e}_1, \dots, \mathbf{e}_n$ ; that is,  $Q(\mathcal{A})(\mathbf{e}_i) = \mathbf{0}$  for all  $i = 1, \dots, n$ . We shall prove that  $Q$  is an annihilator polynomial for every vector  $\mathbf{x} \in L$ . By definition,  $\mathbf{x}$  is a linear combination of vectors of a basis, that is,  $\mathbf{x} = \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \dots + \alpha_n \mathbf{e}_n$ . Then

$$\begin{aligned} Q(\mathcal{A})(\mathbf{x}) &= Q(\mathcal{A})(\alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n) \\ &= \alpha_1 Q(\mathcal{A})(\mathbf{e}_1) + \dots + \alpha_n Q(\mathcal{A})(\mathbf{e}_n) \\ &= \mathbf{0}. \end{aligned}$$

**Definition 4.25** A polynomial that annihilates every vector of a space  $L$  is called an *annihilator polynomial* of this space (keeping in mind that we mean for the given linear transformation  $\mathcal{A} : L \rightarrow L$ ).

In conclusion, let us compare the arguments used in the proofs of Theorems 4.18 and 4.22. In the first case, we relied on the existence of a root (that is, a factor of degree 1) of the *characteristic* polynomial, while in the latter case, we required the existence of a simplest factor (of degree 1 or 2) for the *annihilator* polynomial. The connection between these polynomials relies on a result that is important in and of itself. It is called the *Cayley–Hamilton theorem*.

**Theorem 4.26** *The characteristic polynomial is an annihilator polynomial for its associated vector space.*

The proof of this theorem is based on arguments analogous to those used in the proof of Lemma 4.19, but relating to a much more general situation. We shall now consider polynomials in the variable  $t$  whose coefficients are not numbers, but linear transformations of the vector space  $L$  into itself or (which is the same thing if some fixed basis has been chosen in  $L$ ) square matrices  $P_i$ :

$$P(t) = P_0 + P_1 t + \cdots + P_k t^k.$$

One can work with these as with ordinary polynomials if one assumes that the variable  $t$  commutes with the coefficients. It is also possible to substitute for  $t$  the matrix  $A$  of a linear transformation. We shall denote the result of this substitution by  $P(A)$ , that is,

$$P(A) = P_0 + P_1 A + \cdots + P_k A^k.$$

It is important here that  $t$  and  $A$  are written to the right of the coefficients  $P_i$ . Further, we shall consider the situation in which  $P_i$  and  $A$  are square matrices of one and the same order. In view of what we have said above, all assertions will be true as well for the case that in the last formula, instead of the matrices  $P_i$  and  $A$  we have the linear transformations  $\mathcal{P}_i$  and  $\mathcal{A}$  of some vector space  $L$  into itself:

$$\mathcal{P}(\mathcal{A}) = \mathcal{P}_0 + \mathcal{P}_1 \mathcal{A} + \cdots + \mathcal{P}_k \mathcal{A}^k.$$

However, in this case, the analogue of formula (4.13) from Sect. 4.1 does not hold, that is, if the polynomial  $R(t)$  is equal to  $P(t)Q(t)$  and  $A$  is the matrix of an arbitrary linear transformation of the vector space  $L$ . Then generally speaking,  $R(A) \neq P(A)Q(A)$ . For example, if we have polynomials  $P = P_1 t$  and  $Q = Q_0$ , then  $P_1 t Q_0 = P_1 Q_0 t$ , but it is not true that  $P_1 A Q_0 = P_1 Q_0 A$  for an arbitrary matrix  $A$ , since matrices  $A$  and  $Q_0$  do not necessarily commute. However, there is one important special case in which formula (4.13) holds.

**Lemma 4.27** *Let*

$$P(t) = P_0 + P_1 t + \cdots + P_k t^k, \quad Q(t) = Q_0 + Q_1 t + \cdots + Q_l t^l,$$

*and suppose that the polynomial  $R(t)$  equals  $P(t)Q(t)$ . Then  $R(A) = P(A)Q(A)$  if the matrix  $A$  commutes with every coefficient of the polynomial  $Q(t)$ , that is,  $AQ_i = Q_i A$  for all  $i = 1, \dots, l$ .*

*Proof* It is not difficult to see that the polynomial  $R(t) = P(t)Q(t)$  can be represented in the form  $R(t) = R_0 + R_1 t + \cdots + R_{k+l} t^{k+l}$  with coefficients  $R_s = \sum_{i=0}^s P_i Q_{s-i}$ , where  $P_i = 0$  if  $i > k$ , and  $Q_i = 0$  if  $i > l$ . Similarly, the polynomial  $R(A) = P(A)Q(A)$  can be expressed in the form

$$R(A) = \sum_{s=0}^{k+l} \left( \sum_{i=0}^s P_i A^i Q_{s-i} A^{s-i} \right)$$

with the same conditions:  $P_i = 0$  if  $i > k$ , and  $Q_i = 0$  if  $i > l$ . By the condition of the lemma,  $AQ_j = Q_j A$ , whence by induction, we easily obtain that  $A^i Q_j = Q_j A^i$  for every choice of  $i$  and  $j$ . Thus our expression takes the form

$$R(A) = \sum_{s=0}^{k+l} \left( \sum_{i=0}^s P_i Q_{s-i} A^s \right) = P(A)Q(A). \quad \square$$

Of course, the analogous assertion holds for all polynomials for which the variable  $t$  stands to the left of the coefficients (then the matrix  $A$  must commute with every coefficient of the polynomial  $P$ , and not  $Q$ ).

Using Lemma 4.27, we can prove the Cayley–Hamilton theorem.

*Proof of Theorem 4.26* Let us consider the matrix  $tE - A$  and denote its determinant by  $\varphi(t) = |tE - A|$ . The coefficients of the polynomial  $\varphi(t)$  are numbers, and as is easily seen, it is equal to the characteristic polynomial matrix  $A$  multiplied by  $(-1)^n$  (in order to make the coefficient of  $t^n$  equal to 1). Let us denote by  $B(t)$  the adjugate matrix to  $tE - A$  (see the definition on p. 73). It is clear that  $B(t)$  will contain as its elements certain polynomials in  $t$  of degree at most  $n - 1$ , and consequently, we may write it in the form  $B(t) = B_0 + B_1t + \cdots + B_{n-1}t^{n-1}$ , where the  $B_i$  are certain matrices. Formula (2.70) for the adjugate matrix yields

$$B(t)(tE - A) = \varphi(t)E. \quad (4.22)$$

Let us substitute into formula (4.22) in place of the variable  $t$  the matrix  $A$  of the linear transformation  $\mathcal{A}$  with respect to some basis of the vector space  $L$ . Since the matrix  $A$  commutes with the identity matrix  $E$  and with itself, then by Lemma 4.27, we obtain the matrix equality  $B(A)(AE - A) = \varphi(A)E$ , the left-hand side of which is equal to the null matrix. It is clear that in an arbitrary basis, the null matrix is the matrix of the null transformation  $\mathcal{O} : L \rightarrow L$ , and consequently,  $\varphi(\mathcal{A}) = \mathcal{O}$ . And this is the assertion of Theorem 4.26.  $\square$

In particular, it is now clear that by the proof of Theorem 4.22, we may take as the annihilator polynomial the characteristic polynomial of the transformation  $\mathcal{A}$ .

### 4.3 Complexification

In view of the fact that real vector spaces are encountered especially frequently in applications, we present here another method of determining the properties of linear transformations of such spaces, proceeding from already proved properties of linear transformations of complex spaces.

Let  $L$  be a finite-dimensional real vector space. In order to apply our previously worked-out arguments, it will be necessary to *embed* it in some complex space  $L^{\mathbb{C}}$ . For this, we shall use the fact that, as we saw in Sect. 3.5,  $L$  is isomorphic to the space of rows of length  $n$  (where  $n = \dim L$ ), which we denote by  $\mathbb{R}^n$ .

In view of the usual set inclusion  $\mathbb{R} \subset \mathbb{C}$ , we may consider  $\mathbb{R}^n$  a subset of  $\mathbb{C}^n$ . In this case, it is not, of course, a subspace of  $\mathbb{C}^n$  as a vector space over the field  $\mathbb{C}$ . For example, multiplication by the complex scalar  $i$  does not take  $\mathbb{R}^n$  into itself. On the contrary, as is easily seen, we have the decomposition

$$\mathbb{C}^n = \mathbb{R}^n \oplus i\mathbb{R}^n$$

(let us recall that in  $\mathbb{C}^n$ , multiplication by  $i$  is defined for all vectors, and in particular for vectors in the subset  $\mathbb{R}^n$ ). We shall now denote  $\mathbb{R}^n$  by  $L$ , while  $\mathbb{C}^n$  will be denoted by  $L^{\mathbb{C}}$ . The previous relationship is now written thus:

$$L^{\mathbb{C}} = L \oplus iL. \quad (4.23)$$

An arbitrary linear transformation  $\mathcal{A}$  on a vector space  $L$  (as a space over the field  $\mathbb{R}$ ) can then be extended to all of  $L^{\mathbb{C}}$  (as a space over the field  $\mathbb{C}$ ). Namely, as follows from the decomposition (4.23), every vector  $x \in L^{\mathbb{C}}$  can be uniquely represented in the form  $x = u + iv$ , where  $u, v \in L$ , and we set

$$\mathcal{A}^{\mathbb{C}}(x) = \mathcal{A}(u) + i\mathcal{A}(v). \quad (4.24)$$

We omit the obvious verification that the mapping  $\mathcal{A}^{\mathbb{C}}$  defined by the relationship (4.24) is a linear transformation of the space  $L^{\mathbb{C}}$  (over the field  $\mathbb{C}$ ). Moreover, it is not difficult to prove that  $\mathcal{A}^{\mathbb{C}}$  is the only linear transformation of the space  $L^{\mathbb{C}}$  whose restriction to  $L$  coincides with  $\mathcal{A}$ , that is, for which the equality  $\mathcal{A}^{\mathbb{C}}(x) = \mathcal{A}(x)$  is satisfied for all  $x$  in  $L$ .

The construction presented here may seem somewhat inelegant, since it uses an isomorphism of the spaces  $L$  and  $\mathbb{R}^n$ , for whose construction it is necessary to choose some basis of  $L$ . Although in the majority of applications such a basis exists, we shall give a construction that does not depend on the choice of basis. For this, we recall that the space  $L$  can be reconstructed from its dual space  $L^*$  via the isomorphism  $L \simeq L^{**}$ , which we constructed in Sect. 3.7. In other words,  $L \simeq \mathfrak{L}(L^*, \mathbb{R})$ , where as before,  $\mathfrak{L}(L, M)$  denotes the space of linear mappings  $L \rightarrow M$  (here either all spaces are considered complex or else they are all considered real).

We now consider  $\mathbb{C}$  as a two-dimensional vector space over the field  $\mathbb{R}$  and set

$$L^{\mathbb{C}} = \mathfrak{L}(L^*, \mathbb{C}), \quad (4.25)$$

where in  $\mathfrak{L}(L^*, \mathbb{C})$ , both spaces  $L^*$  and  $\mathbb{C}$  are considered real. Thus the relationship (4.25) carries  $L^{\mathbb{C}}$  into a vector space over the field  $\mathbb{R}$ . But we can convert it into a space over the field  $\mathbb{C}$  after defining multiplication of vectors in  $L^{\mathbb{C}}$  by complex scalars. Namely, if  $\varphi \in \mathfrak{L}(L^*, \mathbb{C})$  and  $z \in \mathbb{C}$ , then we set  $z\varphi = \psi$ , where  $\psi \in \mathfrak{L}(L^*, \mathbb{C})$  is defined by the condition

$$\psi(f) = z \cdot \varphi(f) \quad \text{for all } f \in L^*.$$

It is easily verified that  $L^{\mathbb{C}}$  thus defined is a vector space over the field  $\mathbb{C}$ , and passage from  $L$  to  $L^{\mathbb{C}}$  will be the same as described above, for an arbitrary choice of basis  $L$  (that is, choice of the isomorphism  $L \simeq \mathbb{R}^n$ ).

If  $\mathcal{A}$  is a linear transformation of the space  $L$ , then we shall define a corresponding linear transformation  $\mathcal{A}^{\mathbb{C}}$  of the space  $L^{\mathbb{C}}$ , after assigning to each vector  $\psi \in L^{\mathbb{C}}$  the value  $\mathcal{A}^{\mathbb{C}}(\psi) \in L^{\mathbb{C}}$  using the relation

$$(\mathcal{A}^{\mathbb{C}}(\psi))(f) = \psi(\mathcal{A}^*(f)) \quad \text{for all } f \in L^*,$$

where  $\mathcal{A}^* : L^* \rightarrow L^*$  is the dual transformation to  $\mathcal{A}$  (see p. 125). It is clear that  $\mathcal{A}^{\mathbb{C}}$  is indeed a linear transformation of the space  $L^{\mathbb{C}}$ , and its restriction to  $L$  coincides with the transformation  $\mathcal{A}$ , that is, for every  $\psi \in L$ ,  $\mathcal{A}^{\mathbb{C}}(\psi)(f) = \mathcal{A}(\psi)(f)$  is satisfied for all  $f \in L^*$ .

**Definition 4.28** The complex vector space  $L^{\mathbb{C}}$  is called the *complexification* of the real vector space  $L$ , while the transformation  $\mathcal{A}^{\mathbb{C}} : L^{\mathbb{C}} \rightarrow L^{\mathbb{C}}$  is the *complexification* of the transformation  $\mathcal{A} : L \rightarrow L$ .

*Remark 4.29* The construction presented above is applicable as well to a more general situation: using it, it is possible to assign to any vector space  $L$  over an arbitrary field  $\mathbb{K}$  the space  $L^{\mathbb{K}'}$  over the bigger field  $\mathbb{K}' \supset \mathbb{K}$ , and to the linear transformation  $\mathcal{A}$  of the field  $L$ , the linear transformation  $\mathcal{A}^{\mathbb{K}'}$  of the field  $L^{\mathbb{K}'}$ .

In the space  $L^{\mathbb{C}}$  that we constructed, it will be useful to introduce the operation of complex conjugation, which assigns to a vector  $x \in L^{\mathbb{C}}$  the vector  $\bar{x} \in L^{\mathbb{C}}$ , or interpreting  $L^{\mathbb{C}}$  as  $\mathbb{C}^n$  (with which we began this section), taking the complex conjugate for each number in the row  $x$ , or (equivalently) using (4.23), setting  $\bar{x} = u - iv$  for  $x = u + iv$ . It is clear that

$$\overline{x + y} = \bar{x} + \bar{y}, \quad \overline{(\alpha x)} = \bar{\alpha} \bar{x}$$

hold for all vectors  $x, y \in L^{\mathbb{C}}$  and arbitrary complex scalar  $\alpha$ .

The transformation  $\mathcal{A}^{\mathbb{C}}$  obtained according to the rule (4.24) from a certain transformation  $\mathcal{A}$  of a real vector space  $L$  will be called *real*. For a real transformation  $\mathcal{A}^{\mathbb{C}}$ , we have the relationship

$$\overline{\mathcal{A}^{\mathbb{C}}(x)} = \mathcal{A}^{\mathbb{C}}(\bar{x}), \quad (4.26)$$

which follows from the definition (4.24) of a transformation  $\mathcal{A}^{\mathbb{C}}$ . Indeed, if we have  $x = u + iv$ , then

$$\mathcal{A}^{\mathbb{C}}(x) = \mathcal{A}(u) + i\mathcal{A}(v), \quad \overline{\mathcal{A}^{\mathbb{C}}(x)} = \mathcal{A}(u) - i\mathcal{A}(v).$$

On the other hand,  $\bar{x} = u - iv$ , from which follows  $\mathcal{A}^{\mathbb{C}}(\bar{x}) = \mathcal{A}(u) - i\mathcal{A}(v)$  and therefore (4.26).

Consider the linear transformation  $\mathcal{A}$  of the real vector space  $L$ . To it there corresponds, as shown above, the linear transformation  $\mathcal{A}^{\mathbb{C}}$  of the complex vector space  $L^{\mathbb{C}}$ . By Theorem 4.18, the transformation  $\mathcal{A}^{\mathbb{C}}$  has an eigenvector  $x \in L^{\mathbb{C}}$  for which, therefore, one has the equality

$$\mathcal{A}^{\mathbb{C}}(x) = \lambda x, \quad (4.27)$$

where  $\lambda$  is a root of the characteristic polynomial of the transformation  $\mathcal{A}$  and, generally speaking, is a certain complex number. We must distinguish two cases:  $\lambda$  real and  $\lambda$  complex.



*Case 1:*  $\lambda$  is a real number. In this case, the characteristic polynomial of the transformation  $\mathcal{A}$  has a real root, and therefore  $\mathcal{A}$  has an eigenvector in the field  $L$ ; that is,  $L$  has a one-dimensional invariant subspace.

*Case 2:*  $\lambda$  is a complex number. Let  $\lambda = a + ib$ , where  $a$  and  $b$  are real numbers,  $b \neq 0$ . The eigenvector  $\mathbf{x}$  can also be written in the form  $\mathbf{x} = \mathbf{u} + i\mathbf{v}$ , where the vectors  $\mathbf{u}, \mathbf{v}$  are in  $L$ . By assumption,  $\mathcal{A}^{\mathbb{C}}(\mathbf{x}) = \mathcal{A}(\mathbf{u}) + i\mathcal{A}(\mathbf{v})$ , and then relationship (4.27), in view of the decomposition (4.23), gives

$$\mathcal{A}(\mathbf{v}) = a\mathbf{v} + b\mathbf{u}, \quad \mathcal{A}(\mathbf{u}) = -b\mathbf{v} + a\mathbf{u}. \quad (4.28)$$

This means that the subspace  $L' = \langle \mathbf{v}, \mathbf{u} \rangle$  of the space  $L$  is invariant with respect to the transformation  $\mathcal{A}$ . The dimension of the subspace  $L'$  is equal to 2, and vectors  $\mathbf{v}, \mathbf{u}$  form a basis of it. Indeed, it suffices to verify their linear independence. The linear dependence of  $\mathbf{v}$  and  $\mathbf{u}$  would imply that  $\mathbf{v} = \xi\mathbf{u}$  (or else that  $\mathbf{u} = \xi\mathbf{v}$ ) for some real  $\xi$ . But by  $\mathbf{v} = \xi\mathbf{u}$ , the second equality of (4.28) would yield the relationship  $\mathcal{A}(\mathbf{u}) = (a - b\xi)\mathbf{u}$ , and that would imply that  $\mathbf{u}$  is a real eigenvector of the transformation  $\mathcal{A}$ , with the real eigenvalue  $a - b\xi$ ; that is, we are dealing with case 1. The case  $\mathbf{u} = \xi\mathbf{v}$  is similar.

Uniting cases 1 and 2, we obtain another proof of Theorem 4.22. We observe that in fact, we have now proved even more than what is asserted in that theorem. Namely, we have shown that in the two-dimensional invariant subspace  $L'$  there exists a basis  $\mathbf{v}, \mathbf{u}$  in which the transformation  $\mathcal{A}$  gives the formula (4.28), that is, it has a matrix of the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}, \quad b \neq 0.$$

**Definition 4.30** A linear transformation  $\mathcal{A}$  of a real vector space  $L$  is said to be *block-diagonalizable* if in some basis, its matrix has the form

$$A = \begin{pmatrix} \alpha_1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \alpha_r & 0 & \ddots & \vdots \\ \vdots & \ddots & 0 & B_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & B_s \end{pmatrix}, \quad (4.29)$$

where  $\alpha_1, \dots, \alpha_r$  are real matrices of order 1 (that is, real numbers), and  $B_1, \dots, B_s$  are real matrices of order 2 of the form

$$B_j = \begin{pmatrix} a_j & -b_j \\ b_j & a_j \end{pmatrix}, \quad b_j \neq 0. \quad (4.30)$$

Block-diagonalizable linear transformations are the real analogue of diagonalizable transformations of complex vector spaces. The connection between these two concepts is established in the following theorem.

**Theorem 4.31** *A linear transformation  $\mathcal{A}$  of a vector space  $L$  is block-diagonalizable if and only if its complexification  $\mathcal{A}^{\mathbb{C}}$  is a diagonalizable transformation of the space  $L^{\mathbb{C}}$ .*

*Proof* Suppose the linear transformation  $\mathcal{A} : L \rightarrow L$  is block-diagonalizable. This means that in some basis of the space  $L$ , its matrix has the form (4.29), which is equivalent to the decomposition

$$L = L_1 \oplus \cdots \oplus L_r \oplus M_1 \oplus \cdots \oplus M_s, \quad (4.31)$$

where  $L_i$  and  $M_j$  are subspaces that are invariant with respect to the transformation  $\mathcal{A}$ . In our case,  $\dim L_i = 1$ , so that  $L_i = \langle e_i \rangle$  and  $\mathcal{A}(e_i) = \alpha_i e_i$ , and  $\dim M_j = 2$ , where in some basis of the subspace  $M_j$ , the restriction of the transformation  $\mathcal{A}$  to  $M_j$  has matrix of the form (4.30). Using formula (4.30), one is easily convinced that the restriction  $\mathcal{A}^{\mathbb{C}}$  to the two-dimensional subspace  $M_j$  has two distinct complex-conjugate eigenvalues:  $\lambda_j$  and  $\bar{\lambda}_j$ . If  $f_j$  and  $f'_j$  are the corresponding eigenvectors, then in  $L^{\mathbb{C}}$  there is a basis  $e_1, \dots, e_r, f_1, f'_1, \dots, f_s, f'_s$ , in which the matrix of the transformation  $\mathcal{A}^{\mathbb{C}}$  assumes the form

$$\begin{pmatrix} \alpha_1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \alpha_r & 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & \lambda_1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \bar{\lambda}_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \lambda_s & 0 \\ 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 & \bar{\lambda}_s \end{pmatrix}. \quad (4.32)$$

This means that the transformation  $\mathcal{A}^{\mathbb{C}}$  is diagonalizable.

Now suppose, conversely, that  $\mathcal{A}^{\mathbb{C}}$  is diagonalizable, that is, in some basis of the space  $L^{\mathbb{C}}$ , the transformation  $\mathcal{A}^{\mathbb{C}}$  has the diagonal matrix

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}. \quad (4.33)$$

Among the numbers  $\lambda_1, \dots, \lambda_n$  may be found some that are real and some that are complex. All the numbers  $\lambda_i$  are roots of the characteristic polynomial of the trans-

formation  $\mathcal{A}^{\mathbb{C}}$ . But clearly (by the definition of  $\mathbb{L}^{\mathbb{C}}$ ), any basis of the real vector space  $\mathbb{L}$  is a basis of the complex space  $\mathbb{L}^{\mathbb{C}}$ , and in such a basis, the matrices of the transformations  $\mathcal{A}$  and  $\mathcal{A}^{\mathbb{C}}$  coincide. That is, the matrix of the transformation  $\mathcal{A}^{\mathbb{C}}$  is real in some basis. This means that its characteristic polynomial has real coefficients. It then follows from well-known properties of real polynomials that if among the numbers  $\lambda_1, \dots, \lambda_n$  some are complex, then they come in conjugate pairs  $\lambda_j$  and  $\bar{\lambda}_j$ , and moreover,  $\lambda_j$  and  $\bar{\lambda}_j$  occur the same number of times. We may assume that in the matrix of (4.33), the first  $r$  numbers are real:  $\lambda_i = \alpha_i \in \mathbb{R}$  ( $i \leq r$ ), while the remainder are complex, and moreover,  $\lambda_j$  and  $\bar{\lambda}_j$  ( $j > r$ ) are adjacent to each other. In this case, the matrix of the transformation assumes the form (4.32). Along with each eigenvector  $\mathbf{e}$  of the transformation  $\mathcal{A}^{\mathbb{C}}$ , the space  $\mathbb{L}^{\mathbb{C}}$  contains a vector  $\bar{\mathbf{e}}$ . Moreover, if  $\mathbf{e}$  has the eigenvalue  $\lambda$ , then  $\bar{\mathbf{e}}$  has the eigenvalue  $\bar{\lambda}$ . This follows easily from the fact that  $\mathcal{A}$  is a real transformation and from the relationship  $\overline{(\mathbb{L}^{\mathbb{C}})_{\lambda}} = (\mathbb{L}^{\mathbb{C}})_{\bar{\lambda}}$ , which can be easily verified. Therefore, we may write down the basis in which the transformation  $\mathcal{A}^{\mathbb{C}}$  has the form (4.32) in the form  $\mathbf{e}_1, \dots, \mathbf{e}_r, \mathbf{f}_1, \bar{\mathbf{f}}_1, \dots, \mathbf{f}_s, \bar{\mathbf{f}}_s$ , where all  $\mathbf{e}_i$  are in  $\mathbb{L}$ .

Let us set  $\mathbf{f}_j = \mathbf{u}_j + i\mathbf{v}_j$ , where  $\mathbf{u}_j, \mathbf{v}_j \in \mathbb{L}$ , and let us consider the subspace  $\mathbb{N}_j = \langle \mathbf{u}_j, \mathbf{v}_j \rangle$ . It is clear that  $\mathbb{N}_j$  is invariant with respect to  $\mathcal{A}$ , and by formula (4.28), the restriction of  $\mathcal{A}$  to the subspace  $\mathbb{N}_j$  gives a transformation that in the basis  $\mathbf{u}_j, \mathbf{v}_j$  has matrix of the form (4.30). We therefore see that

$$\mathbb{L}^{\mathbb{C}} = \langle \mathbf{e}_1 \rangle \oplus \dots \oplus \langle \mathbf{e}_r \rangle \oplus i \langle \mathbf{e}_1 \rangle \oplus \dots \oplus i \langle \mathbf{e}_r \rangle \oplus \mathbb{N}_1 \oplus i\mathbb{N}_1 \oplus \dots \oplus \mathbb{N}_s \oplus i\mathbb{N}_s,$$

from which follows the decomposition

$$\mathbb{L} = \langle \mathbf{e}_1 \rangle \oplus \dots \oplus \langle \mathbf{e}_r \rangle \oplus \mathbb{N}_1 \oplus \dots \oplus \mathbb{N}_s,$$

analogous to (4.31). This shows that the transformation  $\mathcal{A} : \mathbb{L} \rightarrow \mathbb{L}$  is block-diagonalizable.  $\square$

Similarly, using the notion of complexification, it is possible to prove a real analogue of Theorems 4.14, 4.18, and 4.21.

## 4.4 Orientation of a Real Vector Space

The real line has two directions: to the *left* and to the *right* (from an arbitrarily chosen point, taken as the origin). Analogously, in real three-dimensional space, there are two directions for traveling around a point: *clockwise* and *counterclockwise*. We shall consider analogous concepts in an arbitrary real vector space (of finite dimension).

Let  $\mathbf{e}_1, \dots, \mathbf{e}_n$  and  $\mathbf{e}'_1, \dots, \mathbf{e}'_n$  be two bases of a real vector space  $\mathbb{L}$ . Then there exists a linear transformation  $\mathcal{A} : \mathbb{L} \rightarrow \mathbb{L}$  such that

$$\mathcal{A}(\mathbf{e}_i) = \mathbf{e}'_i, \quad i = 1, \dots, n. \quad (4.34)$$

It is clear that for the given pair of bases, there exists only one such linear transformation  $\mathcal{A}$ , and moreover, it is not singular: ( $|\mathcal{A}| \neq 0$ ).

**Definition 4.32** Two bases  $e_1, \dots, e_n$  and  $e'_1, \dots, e'_n$  are said to have the *same orientation* if the transformation  $\mathcal{A}$  satisfying the condition (4.34) is proper ( $|\mathcal{A}| > 0$ ; recall Definition 4.4), and to be *oppositely oriented* if  $\mathcal{A}$  is improper ( $|\mathcal{A}| < 0$ ).

**Theorem 4.33** *The property of having the same orientation induces an equivalence relation on the set of all bases of the vector space  $L$ .*

*Proof* The definition of equivalence relation (on an arbitrary set) was given on page xii, and to prove the theorem, we have only to verify symmetry and transitivity, since reflexivity is completely obvious (for the mapping  $\mathcal{A}$ , take the identity transformation  $\mathcal{E}$ ). Since the transformation  $\mathcal{A}$  is nonsingular, it follows that relationship (4.34) can be written in the form  $\mathcal{A}^{-1}(e'_i) = e_i$ ,  $i = 1, \dots, n$ , from which follows the symmetry property of bases having the same orientation: the transformation  $\mathcal{A}$  is replaced by  $\mathcal{A}^{-1}$ , where here  $|\mathcal{A}^{-1}| = |\mathcal{A}|^{-1}$ , and the sign of the determinant remains the same.

Let bases  $e_1, \dots, e_n$  and  $e'_1, \dots, e'_n$  have the same orientation, and suppose bases  $e'_1, \dots, e'_n$  and  $e''_1, \dots, e''_n$  also have the same orientation. By definition, this means that the transformations  $\mathcal{A}$ , from (4.34), and  $\mathcal{B}$ , defined by

$$\mathcal{B}(e'_i) = e''_i, \quad i = 1, \dots, n, \quad (4.35)$$

are proper. Replacing in (4.35) the expressions for the vectors  $e'_i$  from (4.34), we obtain

$$\mathcal{B}\mathcal{A}(e_i) = e''_i, \quad i = 1, \dots, n,$$

and since  $|\mathcal{B}\mathcal{A}| = |\mathcal{B}| \cdot |\mathcal{A}|$ , the transformation  $\mathcal{B}\mathcal{A}$  is also proper, that is, the bases  $e_1, \dots, e_n$  and  $e''_1, \dots, e''_n$  have the same orientation, which completes the proof of transitivity.  $\square$

We shall denote the set of all bases of the space  $L$  by  $\mathfrak{E}$ . Theorem 4.33 then tells us that the property of having the same orientation decomposes the set  $\mathfrak{E}$  into two equivalence classes, that is, we have the decomposition  $\mathfrak{E} = \mathfrak{E}_1 \cup \mathfrak{E}_2$ , where  $\mathfrak{E}_1 \cap \mathfrak{E}_2 = \emptyset$ . To obtain this decomposition in practice, we may proceed as follows: Choose in  $L$  an arbitrary basis  $e_1, \dots, e_n$  and denote by  $\mathfrak{E}_1$  the collection of all bases that have the same orientation as the chosen basis, and let  $\mathfrak{E}_2$  denote the collection of bases with the opposite orientation. Theorem 4.33 tells us that this decomposition of  $\mathfrak{E}$  does not depend on which basis  $e_1, \dots, e_n$  we choose. We can assert that any two bases appearing together in one of the two subsets  $\mathfrak{E}_1$  and  $\mathfrak{E}_2$  have the same orientation, and if they belong to different subsets, then they have opposite orientations.

**Definition 4.34** The choice of one of the subsets  $\mathfrak{E}_1$  and  $\mathfrak{E}_2$  is called an *orientation* of the vector space  $L$ . Once an orientation has been chosen, the bases lying in the

chosen subset are said to be *positively oriented*, while those in the other subset are called *negatively oriented*.

As can be seen from this definition, the selection of an orientation of a vector space depends on an arbitrary choice: it would have been equally possible to have called the positively oriented bases negatively oriented, and vice versa. It is no accident that in practical applications, the actual choice of orientation is frequently based on an appeal such as to the structure of the human body (left–right) or to the motion of the Sun in the heavens (clockwise or counterclockwise).

The crucial part of the theory presented in this section is that there is a connection between orientation and certain topological concepts (such as those presented in the introduction to this book; see p. xvii).

To pursue this idea, we must first of all define *convergence* for sequences of elements of the set  $\mathfrak{E}$ . We shall do so by introducing on the set  $\mathfrak{E}$  a *metric*, that is, by converting it into a *metric space*. This means that we must define a function  $r(x, y)$  for all  $x, y \in \mathfrak{E}$  taking real values and satisfying properties 1–3 introduced on p. xvii. We begin by defining a metric  $r(A, B)$  on the set  $\mathfrak{A}$  of square matrices of a given order  $n$  with real entries.

For a matrix  $A = (a_{ij})$  in  $\mathfrak{A}$ , we let the number  $\mu(A)$  equal the maximum absolute value of its entries:

$$\mu(A) = \max_{i,j=1,\dots,n} |a_{ij}|. \quad (4.36)$$

**Lemma 4.35** *The function  $\mu(A)$  defined by relationship (4.36) exhibits the following properties:*

- (a)  $\mu(A) > 0$  for  $A \neq O$  and  $\mu(A) = 0$  for  $A = O$ .
- (b)  $\mu(A + B) \leq \mu(A) + \mu(B)$  for all  $A, B \in \mathfrak{A}$ .
- (c)  $\mu(AB) \leq n\mu(A)\mu(B)$  for all  $A, B \in \mathfrak{A}$ .

*Proof* Property (a) obviously follows from the definition (4.36), while property (b) follows from an analogous inequality for numbers:  $|a_{ij} + b_{ij}| \leq |a_{ij}| + |b_{ij}|$ . It remains to prove property (c). Let  $A = (a_{ij})$ ,  $B = (b_{ij})$ , and  $C = AB = (c_{ij})$ . Then  $c_{ij} = \sum_{k=1}^n a_{ik}b_{kj}$ , and so

$$|c_{ij}| \leq \sum_{k=1}^n |a_{ik}||b_{kj}| \leq \sum_{k=1}^n \mu(A)\mu(B) = n\mu(A)\mu(B).$$

From this it follows that  $\mu(C) \leq n\mu(A)\mu(B)$ . □

We can now convert the set  $\mathfrak{A}$  into a metric space by setting for every pair of matrices  $A$  and  $B$  in  $\mathfrak{A}$ ,

$$r(A, B) = \mu(A - B). \quad (4.37)$$

Properties 1–3 introduced in the definition of a metric follow from the definitions in (4.36) and (4.37) and properties (a) and (b) proved in Lemma 4.35.

A metric on  $\mathfrak{A}$  enables us to introduce a metric on the set  $\mathfrak{E}$  of bases of a vector space  $L$ . Let us fix a distinguished basis  $e_1, \dots, e_n$  and define the number  $r(x, y)$  for two arbitrary bases  $x$  and  $y$  in the set  $\mathfrak{E}$  as follows. Suppose the bases  $x$  and  $y$  consist of vectors  $x_1, \dots, x_n$  and  $y_1, \dots, y_n$ , respectively. Then there exist linear transformations  $\mathcal{A}$  and  $\mathcal{B}$  of the space  $L$  such that

$$\mathcal{A}(e_i) = x_i, \quad \mathcal{B}(e_i) = y_i, \quad i = 1, \dots, n. \quad (4.38)$$

The transformations  $\mathcal{A}$  and  $\mathcal{B}$  are nonsingular, and by condition (4.38), they are uniquely determined. Let us denote by  $A$  and  $B$  the matrices of the transformations  $\mathcal{A}$  and  $\mathcal{B}$  in the basis  $e_1, \dots, e_n$ , and set

$$r(x, y) = r(A, B), \quad (4.39)$$

where  $r(A, B)$  is as defined above by relationship (4.37). Properties 1–3 in the definition of a metric hold for  $r(x, y)$  from analogous properties of the metric  $r(A, B)$ .

However, here a difficulty arises: The definition of the metric  $r(x, y)$  by relationship (4.39) depends on the choice of some basis  $e_1, \dots, e_n$  of the space  $L$ . Let us choose another basis  $e'_1, \dots, e'_n$  and let us see how the metric  $r'(x, y)$  that results differs from  $r(x, y)$ . To this end, we use the familiar fact that for two bases  $e_1, \dots, e_n$  and  $e'_1, \dots, e'_n$  there exists a unique linear (and in addition, nonsingular) transformation  $\mathcal{C} : L \rightarrow L$  taking the first basis into the second:

$$e'_i = \mathcal{C}(e_i), \quad i = 1, \dots, n. \quad (4.40)$$

Formulas (4.38) and (4.40) show that for linear transformations  $\overline{\mathcal{A}} = \mathcal{A}\mathcal{C}^{-1}$  and  $\overline{\mathcal{B}} = \mathcal{B}\mathcal{C}^{-1}$ , one has the equality

$$\overline{\mathcal{A}}(e'_i) = x_i, \quad \overline{\mathcal{B}}(e'_i) = y_i, \quad i = 1, \dots, n. \quad (4.41)$$

Let us denote by  $A'$  and  $B'$  the matrices of the transformations  $\mathcal{A}$  and  $\mathcal{B}$  in the basis  $e'_1, \dots, e'_n$ , and by  $\overline{A}$  and  $\overline{B}$ , the matrices of the transformations  $\overline{\mathcal{A}}$  and  $\overline{\mathcal{B}}$  in this basis. Let  $C$  be the matrix of the transformation  $\mathcal{C}$ , that is, by (4.40), the transition matrix from the basis  $e'_1, \dots, e'_n$  to the basis  $e_1, \dots, e_n$ . Then matrices  $A', \overline{A}$  and  $B', \overline{B}$  are related by  $\overline{A} = A'C^{-1}$  and  $\overline{B} = B'C^{-1}$ . Furthermore, we observe that  $A$  and  $A'$  are matrices of the same transformation  $\mathcal{A}$  in two different bases ( $e_1, \dots, e_n$  and  $e'_1, \dots, e'_n$ ), and similarly,  $B$  and  $B'$  are matrices of the single transformation  $\mathcal{B}$ . Therefore, by the formula for changing coordinates, we have  $A' = C^{-1}AC$  and  $B' = C^{-1}BC$ , and so as a result, we obtain the relationship

$$\overline{A} = A'C^{-1} = C^{-1}A, \quad \overline{B} = B'C^{-1} = C^{-1}B. \quad (4.42)$$

Returning to the definition (4.39) of a metric on  $\mathfrak{A}$ , we see that  $r'(x, y) = r(\overline{A}, \overline{B})$ . Substituting in the last relationship the expression (4.42) for matrices  $\overline{A}$  and  $\overline{B}$ , and taking into account definition (4.37) and property (c) from Lemma 4.35, we obtain

$$\begin{aligned}
r'(x, y) &= r(\overline{A}, \overline{B}) = r(C^{-1}A, C^{-1}B) \\
&= \mu(C^{-1}(A - B)) \leq n\mu(C^{-1})\mu(A - B) = \alpha r(x, y),
\end{aligned}$$

where the number  $\alpha = n\mu(C^{-1})$  does not depend on the bases  $x$  and  $y$ , but only on  $e_1, \dots, e_n$  and  $e'_1, \dots, e'_n$ . Since the last two bases play a symmetric role in our construction, we may obtain analogously a second equality  $r(x, y) \leq \beta r'(x, y)$  with a certain positive constant  $\beta$ . The relationship

$$r'(x, y) \leq \alpha r(x, y), \quad r(x, y) \leq \beta r'(x, y), \quad \alpha, \quad \beta > 0, \quad (4.43)$$

shows that although the metrics  $r(x, y)$  and  $r'(x, y)$  defined in terms of different bases  $e_1, \dots, e_n$  and  $e'_1, \dots, e'_n$  are different, nevertheless, on the set  $\mathfrak{A}$ , the notion of convergence is the same for both bases. To put this more formally, having chosen in  $\mathfrak{E}$  two different bases and having with the help of these bases defined metrics  $r(x, y)$  and  $r'(x, y)$  on  $\mathfrak{E}$ , we have thereby defined two different metric spaces  $\mathfrak{E}'$  and  $\mathfrak{E}''$  with one and the same underlying set  $\mathfrak{E}$  but with different metrics  $r$  and  $r'$  defined on it. Here the identity mapping of the space  $\mathfrak{E}$  onto itself is not an isometry of  $\mathfrak{E}'$  and  $\mathfrak{E}''$ , but by relationship (4.43), it is a homeomorphism. We may therefore speak about continuous mappings, paths in  $\mathfrak{E}$ , and its connected components without specifying precisely which metric we are using.

Let us move on to the question whether two bases of the set  $\mathfrak{E}$  can be continuously deformed into each other (see the general definition on p. xx). This question reduces to whether there is a continuous deformation between the nonsingular matrices  $A$  and  $B$  corresponding to these bases under the selection of some auxiliary basis  $e_1, \dots, e_n$  (just as with other topological concepts, continuous deformability does not depend on the choice of the auxiliary basis). We wish to emphasize that the condition of nonsingularity of the matrices  $A$  and  $B$  plays here an essential role.

We shall formulate the notion of continuous deformability for matrices in a certain set  $\mathfrak{A}$  (which in our case will be the set of nonsingular matrices).

**Definition 4.36** A matrix  $A$  is said to be *continuously deformable* into a matrix  $B$  if there exists a family of matrices  $A(t)$  in  $\mathfrak{A}$  whose elements depend continuously on a parameter  $t \in [0, 1]$  such that  $A(0) = A$  and  $A(1) = B$ .

It is obvious that this property of matrices being continuously deformable into each other defines an equivalence relation on the set  $\mathfrak{A}$ . By definition, we need to verify that the properties of reflexivity, symmetry, and transitivity are satisfied. The verification of all these properties is simple and given on p. xx.

Let us note one additional property of continuous deformability in the case that the set  $\mathfrak{A}$  has another property: for two arbitrary matrices belonging to  $\mathfrak{A}$ , their product also belongs to  $\mathfrak{A}$ . It is clear that this property is satisfied if  $\mathfrak{A}$  is the set of nonsingular matrices (in subsequent chapters, we shall meet other examples of such sets).

**Lemma 4.37** *If a matrix  $A$  is continuously deformable into  $B$ , and  $C \in \mathfrak{A}$  is an arbitrary matrix, then  $AC$  is continuously deformable into  $BC$ , and  $CA$  is continuously deformable into  $CB$ .*

*Proof* By the condition of the theorem, we have a family  $A(t)$  of matrices in  $\mathfrak{A}$ , where  $t \in [0, 1]$ , effecting a continuous deformation of  $A$  into  $B$ . To prove the first assertion, we take the family  $A(t)C$ , and for the second, the family  $CA(t)$ . This family produces the deformations that we require.  $\square$

**Theorem 4.38** *Two nonsingular square matrices of the same order with real elements are continuously deformable into each other if and only if the signs of their determinants are the same.*

*Proof* Let  $A$  and  $B$  be the matrices described in the statement of the theorem. The necessary condition that the determinants  $|A|$  and  $|B|$  be of the same sign is obvious. Indeed, in view of the formula for the expansion of the determinant (Sect. 2.7) or else by its inductive definition (Sect. 2.2), it is clear that the determinant is a polynomial in the elements of the matrix, and consequently,  $|A(t)|$  is a continuous function of  $t$ . But a continuous function taking values with opposite signs at the endpoints of an interval must take the value zero at some point within the interval, while at the same time, the condition  $|A(t)| \neq 0$  must be satisfied for all  $t \in [0, 1]$ .

Let us prove the sufficiency of the condition, at first for determinants for which  $|A| > 0$ . We shall show that  $A$  is continuously deformable into the identity matrix  $E$ . By Theorem 2.62, the matrix  $A$  can be represented as a product of matrices  $U_{ij}(c)$ ,  $S_k$ , and a diagonal matrix. The matrix  $U_{ij}(c)$  is continuously deformable into the identity: as the family  $A(t)$ , we may take the matrices  $U_{ij}(ct)$ . Since the  $S_k$  are themselves diagonal matrices, we see that (in view of Lemma 4.37) the matrix  $A$  is continuously deformable into the diagonal matrix  $D$ , and from the assumption  $|A| > 0$  and the part of the theorem already proved, it follows that  $|D| > 0$ .

Let

$$D = \begin{pmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & d_3 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & d_n \end{pmatrix}.$$

Every element  $d_i$  can be represented in the form  $\varepsilon_i p_i$ , where  $\varepsilon_i = 1$  or  $-1$ , while  $p_i > 0$ . The matrix  $(p_i)$  of order 1 for  $p_i > 0$  can be continuously deformed into  $(1)$ . For this, it suffices to set  $A(t) = (a(t))$ , where  $a(t) = t + (1-t)p_i$  for  $t \in [0, 1]$ . Therefore, the matrix  $D$  is continuously deformable into the matrix  $D'$ , in which all  $d_i = \varepsilon_i p_i$  are replaced by  $\varepsilon_i$ . As we have seen, from this it follows that  $|D'| > 0$ , that is, the number of  $-1$ 's on the main diagonal is even. Let us combine them in pairs. If there is  $-1$  in the  $i$ th and  $j$ th places, then we recall that the matrix

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \quad (4.44)$$



defines in the plane the central symmetry transformation with respect to the origin, that is, a rotation through the angle  $\pi$ . If we set

$$A(t) = \begin{pmatrix} \cos \pi t & -\sin \pi t \\ \sin \pi t & \cos \pi t \end{pmatrix}, \quad (4.45)$$

then we obtain the matrix of rotation through the angle  $\pi t$ , which as  $t$  changes from 0 to 1, effects a continuous deformation of the matrix (4.44) into the identity. It is clear that we thus obtain a continuous deformation of the matrix  $D'$  into  $E$ .

Denoting continuous deformability by  $\sim$ , we can write down three relationships:  $A \sim D$ ,  $D \sim D'$ ,  $D' \sim E$ , from which follows by transitivity that  $A \sim E$ . From this follows as well the assertion of Theorem 4.38 for two matrices  $A$  and  $B$  with  $|A| > 0$  and  $|B| > 0$ .

In order to take care of matrices  $A$  with  $|A| < 0$ , we introduce the function  $\varepsilon(A) = +1$  if  $|A| > 0$  and  $\varepsilon(A) = -1$  if  $|A| < 0$ . It is clear that  $\varepsilon(AB) = \varepsilon(A)\varepsilon(B)$ . If  $\varepsilon(A) = \varepsilon(B) = -1$ , then let us set  $A^{-1}B = C$ . Then  $\varepsilon(C) = 1$ , and by what was proved previously,  $C \sim E$ . By Lemma 4.37, it follows that  $B \sim A$ , and by symmetry, we have  $A \sim B$ .  $\square$

Taking into account the results of Sect. 3.4 and Lemma 4.37, from Theorem 4.38, we obtain the following result.

**Theorem 4.39** *Two nonsingular linear transformations of a real vector space are continuously deformable into each other if and only if the signs of their determinants are the same.*

**Theorem 4.40** *Two bases of a real vector space are continuously deformable into each other if and only if they have the same orientation.*

Recalling the topological notions introduced earlier of path-connectedness and path-connected component (p. xx), we see that the results we have obtained can be formulated as follows. The set  $\mathfrak{A}$  of nonsingular matrices of a given order (or linear transformations of the space  $L$  into itself) can be represented as the union of two path-connected components corresponding to positive and negative determinants. Similarly, the set  $\mathfrak{E}$  of all bases of a space  $L$  can be represented as the union of two path-connected components consisting of positively and negatively oriented bases.

# Chapter 5

## Jordan Normal Form

### 5.1 Principal Vectors and Cyclic Subspaces

In the previous chapter, we studied linear transformations of real and complex vector spaces into themselves, and in particular, we found conditions under which a linear transformation of a complex vector space is diagonalizable, that is, has a diagonal matrix (consisting of eigenvectors of the transformation) in some specially chosen basis. We showed there that not all transformations of a complex vector space are diagonalizable.

The goal of this chapter is a more complete study of linear transformations of a real or complex vector space to itself, including the investigation of nondiagonalizable transformations. In this chapter as before, we shall denote a vector space by  $L$  and assume that it is finite-dimensional. Moreover, in Sects. 5.1 to 5.3, we shall consider linear transformations of complex vector spaces only.

As already noted, the diagonalizable linear transformations are the simplest class of transformations. However, since this class does not cover all linear transformations, we would like to find a construction that generalizes the construction of diagonalizable linear transformations, and indeed so general as to encompass all linear transformations. A transformation can be brought into diagonal form if there is a basis consisting of the transformation's eigenvectors. Therefore, let us begin by generalizing the notion of eigenvector.

Let us recall that an eigenvector  $e \neq 0$  of a linear transformation  $\mathcal{A} : L \rightarrow L$  with eigenvalue  $\lambda$  satisfies the condition  $\mathcal{A}(e) = \lambda e$ , or equivalently, the equality

$$(\mathcal{A} - \lambda \mathcal{E})(e) = 0.$$

A natural generalization of this is contained in the following definition.

**Definition 5.1** A nonnull vector  $e$  is said to be a *principal vector* of a linear transformation  $\mathcal{A} : L \rightarrow L$  with eigenvalue  $\lambda$  if for some natural number  $m$ , the following condition is satisfied:

$$(\mathcal{A} - \lambda \mathcal{E})^m(e) = 0. \tag{5.1}$$

The smallest natural number  $m$  for which relation (5.1) is satisfied is called the *grade* of the principal vector  $\mathbf{e}$ .

**Example 5.2** An eigenvector is a principal vector of grade 1.

**Example 5.3** Let  $\mathbf{L}$  be the vector space of polynomials  $\mathbf{x}(t)$  of degree at most  $n - 1$ , and let  $\mathcal{A}$  be the linear transformation that maps every function  $\mathbf{x}(t)$  to its derivative  $\mathbf{x}'(t)$ . Then

$$\mathcal{A}(\mathbf{x}(t)) = \mathbf{x}'(t), \quad \mathcal{A}^k(\mathbf{x}(t)) = \mathbf{x}^{(k)}(t).$$

Since  $(t^k)^{(k)} = k! \neq 0$  and  $(t^k)^{(k+1)} = 0$ , it is obvious that the polynomial  $\mathbf{x}(t) = t^k$  is a principal vector of the transformation  $\mathcal{A}$  of grade  $k + 1$  corresponding to the eigenvalue  $\lambda = 0$ .

**Definition 5.4** Let  $\mathbf{e}$  be a principal vector of grade  $m$  corresponding to the eigenvalue  $\lambda$ . The subspace  $\mathbf{M}$  spanned by the vectors

$$\mathbf{e}, \quad (\mathcal{A} - \lambda\mathcal{E})(\mathbf{e}), \quad \dots, \quad (\mathcal{A} - \lambda\mathcal{E})^{m-1}(\mathbf{e}), \quad (5.2)$$

is called the *cyclic subspace* generated by the vector  $\mathbf{e}$ .

**Example 5.5** If  $m = 1$ , then a cyclic subspace is the one-dimensional subspace  $\langle \mathbf{e} \rangle$  generated by the eigenvector  $\mathbf{e}$ .

**Example 5.6** In Example 5.3, the cyclic subspace generated by the principal vector  $\mathbf{x}(t) = t^k$  consists of all polynomials of degree at most  $k$ .

**Theorem 5.7** A cyclic subspace  $\mathbf{M} \subset \mathbf{L}$  generated by the principal vector  $\mathbf{e}$  of grade  $m$  is invariant under the transformation  $\mathcal{A}$  and has dimension  $m$ .

*Proof* Since the cyclic subspace  $\mathbf{M}$  is spanned by  $m$  vectors (5.2), its dimension is obviously at most  $m$ . We shall prove that the vectors (5.2) are linearly independent, which will imply that  $\dim \mathbf{M} = m$ .

Let

$$\alpha_1 \mathbf{e} + \alpha_2 (\mathcal{A} - \lambda\mathcal{E})(\mathbf{e}) + \dots + \alpha_m (\mathcal{A} - \lambda\mathcal{E})^{m-1}(\mathbf{e}) = \mathbf{0}. \quad (5.3)$$

Let us apply the linear transformation  $(\mathcal{A} - \lambda\mathcal{E})^{m-1}$  to both sides of this equality. Since by definition (5.1) of a principal vector, we have  $(\mathcal{A} - \lambda\mathcal{E})^m(\mathbf{e}) = \mathbf{0}$ , then a fortiori,  $(\mathcal{A} - \lambda\mathcal{E})^k(\mathbf{e}) = \mathbf{0}$  for every  $k > m$ . We therefore obtain that

$$\alpha_1 (\mathcal{A} - \lambda\mathcal{E})^{m-1}(\mathbf{e}) = \mathbf{0},$$

and since  $(\mathcal{A} - \lambda\mathcal{E})^{m-1}(\mathbf{e}) \neq \mathbf{0}$ , in view of the fact that  $\mathbf{e}$  is of grade  $m$ , we have the equality  $\alpha_1 = 0$ . Relationship (5.3) now takes the following form:

$$\alpha_2 (\mathcal{A} - \lambda\mathcal{E})(\mathbf{e}) + \dots + \alpha_m (\mathcal{A} - \lambda\mathcal{E})^{m-1}(\mathbf{e}) = \mathbf{0}. \quad (5.4)$$

Applying the linear transformation  $(\mathcal{A} - \lambda\mathcal{E})^{m-2}$  to both parts of equality (5.4), we prove in exactly the same way that  $\alpha_2 = 0$ . Continuing further in this way, we obtain that in relationship (5.3), all the coefficients  $\alpha_1, \dots, \alpha_m$  are equal to zero. Consequently, the vectors (5.2) are linearly independent, and so we have  $\dim \mathbf{M} = m$ .

We shall now prove the invariance of the cyclic subspace  $\mathbf{M}$  associated with the transformation  $\mathcal{A}$ . Let us set

$$\mathbf{e}_1 = \mathbf{e}, \quad \mathbf{e}_2 = (\mathcal{A} - \lambda\mathcal{E})(\mathbf{e}), \quad \dots, \quad \mathbf{e}_m = (\mathcal{A} - \lambda\mathcal{E})^{m-1}(\mathbf{e}). \quad (5.5)$$

Since all vectors of the subspace  $\mathbf{M}$  can be expressed as linear combinations of the vectors  $\mathbf{e}_1, \dots, \mathbf{e}_m$ , it suffices to prove that the vectors  $\mathcal{A}(\mathbf{e}_1), \dots, \mathcal{A}(\mathbf{e}_m)$  can be expressed as linear combinations of  $\mathbf{e}_1, \dots, \mathbf{e}_m$ . But from relationships (5.1) and (5.5), it is clear that

$$(\mathcal{A} - \lambda\mathcal{E})(\mathbf{e}_1) = \mathbf{e}_2, \quad (\mathcal{A} - \lambda\mathcal{E})(\mathbf{e}_2) = \mathbf{e}_3, \quad \dots, \quad (\mathcal{A} - \lambda\mathcal{E})(\mathbf{e}_m) = \mathbf{0},$$

that is,

$$\mathcal{A}(\mathbf{e}_1) = \lambda\mathbf{e}_1 + \mathbf{e}_2, \quad \mathcal{A}(\mathbf{e}_2) = \lambda\mathbf{e}_2 + \mathbf{e}_3, \quad \dots, \quad \mathcal{A}(\mathbf{e}_m) = \lambda\mathbf{e}_m, \quad (5.6)$$

which establishes the assertion of the theorem.  $\square$

**Corollary 5.8** *The vectors  $\mathbf{e}_1, \dots, \mathbf{e}_m$  defined by formula (5.5) form a basis of the cyclic subspace  $\mathbf{M}$  generated by the principal vector  $\mathbf{e}$ . The matrix of the restriction of the linear transformation  $\mathcal{A}$  to the subspace  $\mathbf{M}$  in this basis has the form*

$$A = \begin{pmatrix} \lambda & 0 & 0 & \dots & \dots & 0 \\ 1 & \lambda & 0 & & & 0 \\ 0 & 1 & \lambda & & & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \lambda & 0 \\ 0 & 0 & \dots & \dots & 1 & \lambda \end{pmatrix}. \quad (5.7)$$

This is an obvious consequence of (5.6).

**Theorem 5.9** *Let  $\mathbf{M}$  be a cyclic subspace generated by the principal vector  $\mathbf{e}$  of grade  $m$  with eigenvalue  $\lambda$ . Then an arbitrary vector  $\mathbf{y} \in \mathbf{M}$  can be written in the form*

$$\mathbf{y} = f(\mathcal{A})(\mathbf{e}),$$

where  $f$  is a polynomial of degree at most  $m - 1$ . If the polynomial  $f(t)$  is not divisible by  $t - \lambda$ , then the vector  $\mathbf{y}$  is also a principal vector of grade  $m$  and generates the same cyclic subspace  $\mathbf{M}$ .

*Proof* The first assertion of the theorem follows at once from the fact that by the definition of a cyclic subspace, every vector  $\mathbf{y} \in \mathbf{M}$  has the form

$$\mathbf{y} = \alpha_1 \mathbf{e} + \alpha_2 (\mathcal{A} - \lambda \mathcal{E})(\mathbf{e}) + \cdots + \alpha_m (\mathcal{A} - \lambda \mathcal{E})^{m-1}(\mathbf{e}), \quad (5.8)$$

that is,  $\mathbf{y} = f(\mathcal{A})(\mathbf{e})$ , where the polynomial  $f(t)$  is given by

$$f(t) = \alpha_1 + \alpha_2(t - \lambda) + \cdots + \alpha_m(t - \lambda)^{m-1}.$$

Let us prove the second assertion. Let  $\mathbf{y} = f(\mathcal{A})(\mathbf{e})$ . Then  $(\mathcal{A} - \lambda \mathcal{E})^m(\mathbf{y}) = \mathbf{0}$ . Indeed, from the relationships  $\mathbf{y} = f(\mathcal{A})(\mathbf{e})$  and (5.1) and taking into account the property established earlier that two arbitrary polynomials in one and the same linear transformation commute (a consequence of Lemma 4.16 in Sect. 4.1; see p. 142), we obtain the equality

$$(\mathcal{A} - \lambda \mathcal{E})^m(\mathbf{y}) = (\mathcal{A} - \lambda \mathcal{E})^m f(\mathcal{A})(\mathbf{e}) = f(\mathcal{A})(\mathcal{A} - \lambda \mathcal{E})^m(\mathbf{e}) = \mathbf{0}.$$

Let us assume that the polynomial  $f(t)$  is not divisible by  $t - \lambda$ . This implies that the coefficient  $\alpha_1$  is nonzero. We shall show that we then must have  $(\mathcal{A} - \lambda \mathcal{E})^{m-1}(\mathbf{y}) \neq \mathbf{0}$ . Applying the linear transformation  $(\mathcal{A} - \lambda \mathcal{E})^{m-1}$  to the vectors on both sides of equality (5.8), we obtain

$$\begin{aligned} (\mathcal{A} - \lambda \mathcal{E})^{m-1}(\mathbf{y}) &= \alpha_1 (\mathcal{A} - \lambda \mathcal{E})^{m-1}(\mathbf{e}) + \alpha_2 (\mathcal{A} - \lambda \mathcal{E})^m(\mathbf{e}) + \cdots + \alpha_m (\mathcal{A} - \lambda \mathcal{E})^{2m-2}(\mathbf{e}) \\ &= \alpha_1 (\mathcal{A} - \lambda \mathcal{E})^{m-1}(\mathbf{e}), \end{aligned}$$

since we have  $(\mathcal{A} - \lambda \mathcal{E})^k(\mathbf{e}) = \mathbf{0}$  for every  $k \geq m$ . From this last relationship and taking into account the conditions  $\alpha_1 \neq 0$  and  $(\mathcal{A} - \lambda \mathcal{E})^{m-1}(\mathbf{e}) \neq \mathbf{0}$ , it follows that  $(\mathcal{A} - \lambda \mathcal{E})^{m-1}(\mathbf{y}) \neq \mathbf{0}$ . Therefore, the vector  $\mathbf{y}$  is also a principal vector of the linear transformation  $\mathcal{A}$  of grade  $m$ .

Finally, we shall prove that the cyclic subspaces  $\mathbf{M}$  and  $\mathbf{M}'$  generated by principal vectors  $\mathbf{e}$  and  $\mathbf{y}$  coincide. It is clear that  $\mathbf{M}' \subset \mathbf{M}$ , since  $\mathbf{y} \in \mathbf{M}$ , and in view of the invariance of the cyclic subspace  $\mathbf{M}$ , the vector  $(\mathcal{A} - \lambda \mathcal{E})^k(\mathbf{y})$  for arbitrary  $k$  is also contained in  $\mathbf{M}$ . But from Theorem 5.7, it follows that  $\dim \mathbf{M} = \dim \mathbf{M}' = m$ , and therefore, by Theorem 3.24, the inclusion  $\mathbf{M}' \subset \mathbf{M}$  implies simply the equality  $\mathbf{M}' = \mathbf{M}$ .  $\square$

**Corollary 5.10** *In the notation of Theorem 5.9, for an arbitrary vector  $\mathbf{y} \in \mathbf{M}$  and scalar  $\mu \neq \lambda$ , we have the representation  $\mathbf{y} = (\mathcal{A} - \mu \mathcal{E})(\mathbf{z})$  for some vector  $\mathbf{z} \in \mathbf{M}$ . Furthermore, we have the following: either  $\mathbf{y}$  is a principal vector of grade  $m$  that generates the cyclic subspace  $\mathbf{M}$ , or else  $\mathbf{y} = (\mathcal{A} - \lambda \mathcal{E})(\mathbf{z})$  for some vector  $\mathbf{z} \in \mathbf{M}$ .*

*Proof* The matrix of the restriction of the linear transformation  $\mathcal{A}$  to the subspace  $\mathbf{M}$  in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_m$  from (5.5) has the form (5.7). From this, it is easily seen that for arbitrary  $\mu \neq \lambda$ , the determinant of the restriction of the linear transformation

$\mathcal{A} - \mu\mathcal{E}$  to  $M$  is nonzero. From Theorems 3.69 and 3.70, it follows that the restriction of  $\mathcal{A} - \mu\mathcal{E}$  to  $M$  is an isomorphism  $M \xrightarrow{\sim} M$ , and its image is  $(\mathcal{A} - \mu\mathcal{E})(M) = M$ ; that is, for an arbitrary vector  $y \in M$ , there exists a vector  $z \in M$  such that  $y = (\mathcal{A} - \mu\mathcal{E})(z)$ .

By Theorem 5.9, a vector  $y$  can be represented in the form  $y = f(\mathcal{A})(e)$ , and moreover, if the polynomial  $f(t)$  is not divisible by  $t - \lambda$ , then  $y$  is a principal vector of grade  $m$  generating the cyclic subspace  $M$ . But if  $f(t)$  is divisible by  $t - \lambda$ , that is,  $f(t) = (t - \lambda)g(t)$  for some polynomial  $g(t)$ , then setting  $z = g(\mathcal{A})(e)$ , we obtain the required representation  $y = (\mathcal{A} - \lambda\mathcal{E})(z)$ .  $\square$

## 5.2 Jordan Normal Form (Decomposition)

For the proof of the major result of this section and indeed of the entire chapter—the theorem on the decomposition of a complex vector space as a direct sum of cyclic subspaces—we require the following lemma.

**Lemma 5.11** *For an arbitrary linear transformation  $\mathcal{A} : L \rightarrow L$  of a complex vector space, there exist a scalar  $\lambda$  and an  $(n - 1)$ -dimensional subspace  $L' \subset L$  invariant with respect to the transformation  $\mathcal{A}$  such that for every vector  $x \in L$ , we have the equality*

$$\mathcal{A}(x) = \lambda x + y, \quad \text{where } y \in L'. \quad (5.9)$$

*Proof* By Theorem 4.18, every linear transformation of a complex vector space has an eigenvector and associated eigenvalue. Let  $\lambda$  be an eigenvalue of the transformation  $\mathcal{A}$ . Then the transformation  $\mathcal{B} = \mathcal{A} - \lambda\mathcal{E}$  is singular (it annihilates the eigenvector), and by Theorem 3.72, its image  $\mathcal{B}(L)$  is a subspace  $M \subset L$  of dimension  $m < n$ .

Let  $e_1, \dots, e_m$  be a basis of  $M$ . We shall extend it arbitrarily to a basis of  $L$  by means of the vectors  $e_{m+1}, \dots, e_n$ . It is clear that the subspace

$$L' = \langle e_1, \dots, e_m, e_{m+1}, \dots, e_{n-1} \rangle$$

has dimension  $n - 1$  and includes  $M$ , since  $e_1, \dots, e_m \in M$ .

Let us now prove equality (5.9). Consider an arbitrary vector  $x \in L$ . Then we have  $\mathcal{B}(x) \in \mathcal{B}(L) = M$ , which implies that  $\mathcal{B}(x) \in L'$ , since  $M \subset L'$ . Recalling that  $\mathcal{A} = \mathcal{B} + \lambda\mathcal{E}$ , we obtain that  $\mathcal{A}(x) = \mathcal{B}(x) + \lambda x$ , and moreover, by our construction, the vector  $y = \mathcal{B}(x)$  is in  $L'$ . From this, the invariance of the subspace  $L'$  easily follows. Indeed, if  $x \in L'$ , then in equality (5.9), we have not only  $y \in L'$ , but also  $\lambda x \in L'$ , which yields that  $\mathcal{A}(x) \in L'$  as well.  $\square$

The main result of this section (the decomposition theorem) is the following.

**Theorem 5.12** *A finite-dimensional complex vector space  $L$  can be decomposed as a direct sum of cyclic subspaces relative to an arbitrary linear transformation  $\mathcal{A} : L \rightarrow L$ .*

*Proof* The proof will be by induction on the dimension  $n = \dim L$ . It is based on the lemma proved above, and we shall use the same notation. Let  $L' \subset L$  be the same  $(n - 1)$ -dimensional subspace invariant with respect to the transformation  $\mathcal{A}$  that was discussed in Lemma 5.11.

We choose any vector  $e' \notin L'$ . If  $f_1, \dots, f_{n-1}$  is any basis of the subspace  $L'$ , then the vectors  $f_1, \dots, f_{n-1}, e'$  form a basis of  $L$ . Indeed, there are  $n = \dim L$  vectors, and so it suffices to prove their linear independence. Let us suppose that

$$\alpha_1 f_1 + \dots + \alpha_{n-1} f_{n-1} + \beta e' = 0. \quad (5.10)$$

If  $\beta \neq 0$ , then from this equality, it would follow that  $e' \in L'$ . Therefore,  $\beta = 0$ , and then from equality (5.10), by the linear independence of the vectors  $f_1, \dots, f_{n-1}$  it follows that  $\alpha_1 = \dots = \alpha_{n-1} = 0$ .

We shall rely on the fact that the vector  $e' \in L$  can be chosen arbitrarily. Till now, it satisfied only the single condition  $e' \notin L'$ , but it is not difficult to see that every vector  $e'' = e' + x$ , where  $x \in L'$ , satisfies the same condition, and this means that any such vector could have been chosen in place of  $e'$ . Indeed, if  $e'' \in L'$ , then considering that  $x \in L'$ , we would have  $e' \in L'$ , contradicting the assumption.

It is obvious that Theorem 5.12 is true for  $n = 1$ . Therefore, by the induction hypothesis, we may assume that it holds as well for the subspace  $L'$ . Let

$$L' = L_1 \oplus \dots \oplus L_r \quad (5.11)$$

be the decomposition of  $L'$  as a sum of cyclic subspaces, and moreover, suppose that each cyclic subspace  $L_i$  is generated by its principal vector  $e_i$  of grade  $m_i$  associated with the eigenvalue  $\lambda_i$  and has the basis

$$e_i, \quad (\mathcal{A} - \lambda_i \mathcal{E})(e_i), \quad \dots, \quad (\mathcal{A} - \lambda_i \mathcal{E})^{m_i-1}(e_i). \quad (5.12)$$

By Theorem 5.7, it follows that  $\dim L_i = m_i$  and  $n - 1 = m_1 + \dots + m_r$ .

For the vector  $e'$  chosen at the start of the proof, we have, by the lemma, the equality

$$\mathcal{A}(e') = \lambda e' + y, \quad \text{where } y \in L'.$$

In view of the decomposition (5.11), this vector  $y$  can be written in the form

$$y = y_1 + \dots + y_r, \quad (5.13)$$

where  $y_i \in L_i$ . Thanks to Corollary 5.10, we may assert that the vector  $y_i$  either can be written in the form  $(\mathcal{A} - \lambda \mathcal{E})(z_i)$  for some  $z_i \in L_i$ , or is a principal vector of grade  $m_i$  associated with the eigenvalue  $\lambda$ . Changing if necessary the numeration of the vectors  $y_i$ , we may write

$$(\mathcal{A} - \lambda \mathcal{E})(e') = (\mathcal{A} - \lambda \mathcal{E})(z) + y_s + \dots + y_r, \quad (5.14)$$

where  $z = z_1 + \dots + z_{s-1}$ ,  $z_i \in L_i$ , for all  $i = 1, \dots, s - 1$ , and each of the vectors  $y_j$  with indices  $j = s, \dots, r$  generates the cyclic subspace  $L_j$ .

Here there are two possible cases.

*Case 1.* In formula (5.14), we have  $s - 1 = r$ , that is,

$$(\mathcal{A} - \lambda \mathcal{E})(\mathbf{e}') = (\mathcal{A} - \lambda \mathcal{E})(\mathbf{z}), \quad \mathbf{z} \in \mathbf{L}'.$$

Choosing the vector  $\mathbf{e}'$  arbitrarily, as discussed above, we set  $\mathbf{e}'' = \mathbf{e}' - \mathbf{z}$ . Then from the previous relationship, we obtain

$$(\mathcal{A} - \lambda \mathcal{E})(\mathbf{e}'') = \mathbf{0}.$$

By definition, this implies that  $\mathbf{e}''$  is an eigenvector with eigenvalue  $\lambda$ . Consider the one-dimensional subspace  $\mathbf{L}_{r+1} = \langle \mathbf{e}'' \rangle$ . It is clear that it is cyclic, and moreover,

$$\mathbf{L} = \mathbf{L}' \oplus \mathbf{L}_{r+1} = \mathbf{L}_1 \oplus \cdots \oplus \mathbf{L}_r \oplus \mathbf{L}_{r+1}.$$

Theorem 5.12 has been proved in this case.

*Case 2.* In formula (5.14), we have  $s - 1 < r$ . We again set  $\mathbf{e}'' = \mathbf{e}' - \mathbf{z}$ . Then from (5.14), we obtain that

$$(\mathcal{A} - \lambda \mathcal{E})(\mathbf{e}'') = \mathbf{y}_s + \cdots + \mathbf{y}_r, \quad (5.15)$$

where by construction, each  $\mathbf{y}_j$ ,  $j = s, \dots, r$ , is a principal vector of grade  $m_j$  corresponding to the eigenvalue  $\lambda$  generating the cyclic subspace  $\mathbf{L}_j$ .

It is clear that we can always order the vectors  $\mathbf{y}_s, \dots, \mathbf{y}_r$  in such a way that  $m_s \leq \cdots \leq m_r$ . Let us assume that this condition is satisfied. We shall prove that the vector  $\mathbf{e}''$  is a principal vector of grade  $m_r + 1$  with associated eigenvalue  $\lambda$ , and we shall show that we then have the following decomposition:

$$\mathbf{L} = \mathbf{L}_1 \oplus \cdots \oplus \mathbf{L}_{r-1} \oplus \mathbf{L}'_r, \quad (5.16)$$

where  $\mathbf{L}'_r$  is a cyclic subspace generated by the vector  $\mathbf{e}''$ . It is clear that from this will follow the assertion of Theorem 5.12. From the equality (5.15), it follows that

$$(\mathcal{A} - \lambda \mathcal{E})^{m_r+1}(\mathbf{e}'') = (\mathcal{A} - \lambda \mathcal{E})^{m_r}(\mathbf{y}_s) + \cdots + (\mathcal{A} - \lambda \mathcal{E})^{m_r}(\mathbf{y}_r). \quad (5.17)$$

Since the principal vectors  $\mathbf{y}_i$ ,  $i = s, \dots, r$ , have grades  $m_i$ , and since by our assumption, all the  $m_i$  are less than or equal to  $m_r$ , it follows that  $(\mathcal{A} - \lambda \mathcal{E})^{m_r}(\mathbf{y}_i) = \mathbf{0}$  for all  $i = s, \dots, r$ . From this, taking into account (5.17), it follows that  $(\mathcal{A} - \lambda \mathcal{E})^{m_r+1}(\mathbf{e}'') = \mathbf{0}$ . In just the same way, we obtain that

$$(\mathcal{A} - \lambda \mathcal{E})^{m_r}(\mathbf{e}'') = (\mathcal{A} - \lambda \mathcal{E})^{m_r-1}(\mathbf{y}_s) + \cdots + (\mathcal{A} - \lambda \mathcal{E})^{m_r-1}(\mathbf{y}_r). \quad (5.18)$$

The terms on the right-hand side of this sum belong to the subspaces  $\mathbf{L}_s, \dots, \mathbf{L}_r$ . If we had the equality

$$(\mathcal{A} - \lambda \mathcal{E})^{m_r}(\mathbf{e}'') = \mathbf{0},$$



then it would follow that all the terms on the right-hand side of (5.18) would be equal to zero, since the subspaces  $L_s, \dots, L_r$  form a direct sum. In particular, we would obtain that  $(\mathcal{A} - \lambda\mathcal{E})^{m_r-1}(\mathbf{y}_r) = \mathbf{0}$ , and this would contradict that the principal vector  $\mathbf{y}_r$  has grade  $m_r$ . We therefore conclude that  $(\mathcal{A} - \lambda\mathcal{E})^{m_r}(\mathbf{e}'') \neq \mathbf{0}$ , and consequently, the principal vector  $\mathbf{e}''$  has grade  $m_r + 1$ .

It remains to prove relationship (5.16). We observe that the dimensions of the spaces  $L_1, \dots, L_{r-1}$  are equal to  $m_1, \dots, m_{r-1}$ , while the dimension of  $L'_r$  is equal to  $m_r + 1$ . Therefore, from equality (5.12), it follows that the sum of the dimensions of the terms on the right-hand side of (5.16) equals the dimension of the left-hand side. Therefore, in order to prove the relationship (5.16), it suffices by Corollary 3.40 (p. 96) to prove that an arbitrary vector in the space  $L$  can be represented as the sum of vectors from the subspaces  $L_1, \dots, L_{r-1}, L'_r$ .

It suffices to prove this last assertion for all vectors in a certain basis of the space  $L$ . Such a basis is obtained in particular if we combine the vector  $\mathbf{e}''$  and the vectors of certain bases of the subspaces  $L_1, \dots, L_r$ . For the vector  $\mathbf{e}''$ , this assertion is obvious, since  $\mathbf{e}'' \in L'_r$ . In just the same way, the assertion is clear for any vector in the basis of one of the subspaces  $L_1, \dots, L_{r-1}$ . It remains to prove this for vectors in some basis of the subspace  $L_r$ . Such a basis, for example, comprises the vectors

$$\mathbf{y}_r, \quad (\mathcal{A} - \lambda\mathcal{E})(\mathbf{y}_r), \quad \dots, \quad (\mathcal{A} - \lambda\mathcal{E})^{m_r-1}(\mathbf{y}_r).$$

From (5.15), it follows that

$$\mathbf{y}_r = -(\mathbf{y}_s + \dots + \mathbf{y}_{r-1}) + (\mathcal{A} - \lambda\mathcal{E})(\mathbf{e}''),$$

and this means that

$$(\mathcal{A} - \lambda\mathcal{E})^k(\mathbf{y}_r) = -(\mathcal{A} - \lambda\mathcal{E})^k(\mathbf{y}_s) - \dots - (\mathcal{A} - \lambda\mathcal{E})^k(\mathbf{y}_{r-1}) + (\mathcal{A} - \lambda\mathcal{E})^{k+1}(\mathbf{e}'')$$

for all  $k = 1, \dots, m_r - 1$ . And this establishes what we needed to show: since

$$\mathbf{y}_s \in L_s, \quad \dots, \quad \mathbf{y}_{r-1} \in L_{r-1}, \quad \mathbf{e}'' \in L'_r,$$

and since the spaces  $L_s, \dots, L_{r-1}$  and  $L'_r$  are invariant, it follows that

$$\begin{aligned} (\mathcal{A} - \lambda\mathcal{E})^k(\mathbf{y}_s) &\in L_s, & \dots, & & (\mathcal{A} - \lambda\mathcal{E})^k(\mathbf{y}_{r-1}) &\in L_{r-1}, \\ (\mathcal{A} - \lambda\mathcal{E})^{k+1}(\mathbf{e}'') &\in L'_r. \end{aligned}$$

This completes the proof of Theorem 5.12. □

Let us note that in the passage from the subspace  $L'$  to  $L$  for a given  $\lambda$ , the decomposition into cyclic subspaces changes in the following way: either in the decomposition there appears one more one-dimensional subspace (case 1), or else the dimension of one of the cyclic subspaces increases by 1 (case 2).

Let the decomposition into a direct sum of subspaces, whose existence is established by Theorem 5.12, have the form

$$L = L_1 \oplus \dots \oplus L_r.$$

In each of the subspaces  $L_i$ , we will select a basis of the form (5.5) and combine them into a single basis  $e_1, \dots, e_n$  of the space  $L$ . In this basis, the matrix  $A$  of the transformation  $\mathcal{A}$  has the block-diagonal form

$$A = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_r \end{pmatrix}, \quad (5.19)$$

where the matrices  $A_i$  have (by Corollary 5.8) the form

$$A_i = \begin{pmatrix} \lambda_i & 0 & 0 & \cdots & \cdots & 0 \\ 1 & \lambda_i & 0 & & & 0 \\ 0 & 1 & \lambda_i & & & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \lambda_i & 0 \\ 0 & 0 & \cdots & \cdots & 1 & \lambda_i \end{pmatrix}. \quad (5.20)$$

The matrix  $A$  given by formulas (5.19) and (5.20) is said to be in *Jordan normal form*, while the matrices  $A_i$  are called *Jordan blocks*. We therefore have the following result, which is nothing more than a reformulation of Theorem 5.12.

**Theorem 5.13** *For every linear transformation of a finite-dimensional complex vector space, there exists a basis of that space in which the matrix of the transformation is in Jordan normal form.*

**Corollary 5.14** *Every complex matrix is similar to a matrix in Jordan normal form.*

*Proof* As we saw in Chap. 3, an arbitrary square matrix  $A$  of order  $n$  is the matrix of some linear transformation  $\mathcal{A} : L \rightarrow L$  in some basis  $e_1, \dots, e_n$ . By Theorem 5.13, in some other basis  $e'_1, \dots, e'_n$ , the matrix  $A'$  of the transformation  $\mathcal{A}$  is in Jordan normal form. As established in Sect. 3.4, the matrices  $A$  and  $A'$  are related by the relationship (3.43), for some nonsingular matrix  $C$  (the transition matrix from the first basis to the second). This implies that the matrices  $A$  and  $A'$  are similar.  $\square$

### 5.3 Jordan Normal Form (Uniqueness)

We shall now explore the extent to which the decomposition of the vector space  $L$  as a direct sum of cyclic subspaces relative to a given linear transformation  $\mathcal{A} : L \rightarrow L$  is unique. First of all, let us remark that in such a decomposition

$$L = L_1 \oplus \cdots \oplus L_r, \quad (5.21)$$

the subspaces  $L_i$  themselves are in no way uniquely determined. The simplest example of this is the identity transformation  $\mathcal{A} = \mathcal{E}$ . For this transformation, every nonnull vector is an eigenvector, which means that every one-dimensional subspace is a cyclic subspace generated by a principal vector of grade 1. Therefore, any decomposition of the space  $L$  as a direct sum of one-dimensional subspaces is a decomposition as a direct sum of cyclic subspaces, and such a decomposition exists for every basis of the space  $L$ ; that is, there are infinitely many of them.

However, we shall prove that eigenvalues  $\lambda_i$  and the dimensions of the cyclic subspaces associated with these numbers coincide for every possible decomposition (5.21). As we have seen, the Jordan normal form is determined solely by the eigenvalues  $\lambda_i$  and the dimensions of the associated subspaces (see formulas (5.19) and (5.20)). This will give us the uniqueness of the Jordan normal form.

**Theorem 5.15** *The Jordan normal form of a linear transformation is completely determined by the transformation itself up to the ordering of the Jordan blocks. In other words, for the decomposition (5.21) of a vector space  $L$  as a direct sum of subspaces that are cyclic for some linear transformation  $\mathcal{A} : L \rightarrow L$ , the eigenvalues  $\lambda_i$  and dimensions  $m_i$  of the associated cyclic subspaces  $L_i$  depend only on the transformation  $\mathcal{A}$  and are the same for all decompositions (5.21).*

*Proof* Let  $\lambda$  be some eigenvalue of the linear transformation  $\mathcal{A}$  and let (5.21) be one possible decomposition. Let us denote by  $l_m$  ( $m = 1, 2, \dots$ ) the integer that indicates how many  $m$ -dimensional cyclic subspaces associated with  $\lambda$  are encountered in (5.21).

We shall give a method for calculating  $l_m$ , based on  $\lambda$  and  $\mathcal{A}$  only. This will prove that this number in fact does not depend on the decomposition (5.21).

Let us apply to both sides of equality (5.21) the transformation  $(\mathcal{A} - \lambda\mathcal{E})^i$  with some  $i \geq 1$ . It is clear that

$$(\mathcal{A} - \lambda\mathcal{E})^i(L) = (\mathcal{A} - \lambda\mathcal{E})^i(L_1) \oplus \dots \oplus (\mathcal{A} - \lambda\mathcal{E})^i(L_r). \quad (5.22)$$

We shall now determine the dimensions of the subspaces  $(\mathcal{A} - \lambda\mathcal{E})^i(L_k)$ . In the course of proving the corollary to Theorem 5.9 (Corollary 5.10), we established that for arbitrary  $\mu \neq \lambda$ , the restriction of the linear transformation  $\mathcal{A} - \mu\mathcal{E}$  to  $M$  is an isomorphism, and its image  $(\mathcal{A} - \mu\mathcal{E})(M)$  is equal to  $M$ . Therefore, if  $L_k$  corresponds to the number  $\lambda_k \neq \lambda$ , then

$$(\mathcal{A} - \lambda\mathcal{E})^i(L_k) = L_k, \quad \lambda_k \neq \lambda. \quad (5.23)$$

But if  $\lambda_k = \lambda$ , then choosing in  $L_k$  the basis  $\mathbf{e}, (\mathcal{A} - \lambda\mathcal{E})(\mathbf{e}), \dots, (\mathcal{A} - \lambda\mathcal{E})^{m_k-1}(\mathbf{e})$ , where  $m_k = \dim L_k$ , that is, it is equal to the grade of the principal vector  $\mathbf{e}$ , we obtain that if  $i \geq m_k$ , then the subspace  $(\mathcal{A} - \lambda\mathcal{E})^i(L_k)$  consists solely of the null vector, while if  $i < m_k$ , then

$$(\mathcal{A} - \lambda\mathcal{E})^i(L_k) = \langle (\mathcal{A} - \lambda\mathcal{E})^i(\mathbf{e}), \dots, (\mathcal{A} - \lambda\mathcal{E})^{m_k-1}(\mathbf{e}) \rangle,$$

and moreover, the vectors  $(\mathcal{A} - \lambda\mathcal{E})^i(\mathbf{e}), \dots, (\mathcal{A} - \lambda\mathcal{E})^{m_k-1}(\mathbf{e})$  are linearly independent. Therefore, in the case  $\lambda_k = \lambda$ , we obtain the formula

$$\dim(\mathcal{A} - \lambda\mathcal{E})^i(\mathbf{L}_k) = \begin{cases} 0, & \text{if } i \geq m_k, \\ m_k - i, & \text{if } i < m_k. \end{cases} \quad (5.24)$$

Let us denote by  $n'$  the sum of the dimensions of those subspaces  $\mathbf{L}_k$  that correspond to the numbers  $\lambda_k \neq \lambda$ . Then from formulas (5.22)–(5.24), it follows that

$$\dim(\mathcal{A} - \lambda\mathcal{E})^i(\mathbf{L}) = l_{i+1} + 2l_{i+2} + \dots + (p-i)l_p + n', \quad (5.25)$$

where  $p$  is the maximal dimension of a cyclic subspace associated with the given value  $\lambda$  in the decomposition (5.21). Indeed, from the equality (5.22), we obtain that

$$\dim(\mathcal{A} - \lambda\mathcal{E})^i(\mathbf{L}) = \dim(\mathcal{A} - \lambda\mathcal{E})^i(\mathbf{L}_1) + \dots + \dim(\mathcal{A} - \lambda\mathcal{E})^i(\mathbf{L}_r). \quad (5.26)$$

It follows from formula (5.23) that the terms  $\dim(\mathcal{A} - \lambda\mathcal{E})^i(\mathbf{L}_k)$  with  $\lambda_k \neq \lambda$  in the sum give  $n'$ . In view of formula (5.24), the terms  $\dim(\mathcal{A} - \lambda\mathcal{E})^i(\mathbf{L}_k)$  with  $\lambda_k = \lambda$  and  $m_k \leq i$  are equal to zero. Furthermore, from the same formula (5.24), it follows that if  $m_k = i + 1$ , then  $\dim(\mathcal{A} - \lambda\mathcal{E})^i(\mathbf{L}_k) = 1$ , and the number of subspaces  $\mathbf{L}_k$  of dimension  $m_k = i + 1$  will be equal to  $l_{i+1}$  by the definition of the number  $l_m$ . Therefore, in formula (5.26), the number of terms equal to 1 will be  $l_{i+1}$ . Similarly, the number of subspaces  $\mathbf{L}_k$  of dimension  $m_k = i + 2$  will be equal  $l_{i+2}$ , but with this, we already have  $\dim(\mathcal{A} - \lambda\mathcal{E})^i(\mathbf{L}_k) = 2$ , whence on the right-hand side of (5.25), there appears the term  $2l_{i+2}$ , and so on. From this follows the equality (5.25).

Let us recall that in Sect. 3.6, we defined the notion of the rank  $\text{rk } \mathcal{B}$  of an arbitrary linear transformation  $\mathcal{B} : \mathbf{L} \rightarrow \mathbf{L}$ . Here,  $\text{rk } \mathcal{B}$  coincides with the dimension of the image  $\mathcal{B}(\mathbf{L})$  and is equal to the rank of the matrix  $B$  of this transformation, regardless of the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  in terms of which the matrix of the transformation is written.

Let us now set  $r_i = \text{rk}(\mathcal{A} - \lambda\mathcal{E})^i$  for  $i = 1, \dots, p$ . Let us write the relationships (5.25) for  $i = 1, \dots, p$  by taking into account the fact that

$$\dim(\mathcal{A} - \lambda\mathcal{E})^i(\mathbf{L}) = \text{rk}(\mathcal{A} - \lambda\mathcal{E})^i = r_i \quad \text{and} \quad l_s = 0 \quad \text{for } s > p,$$

and let us consider also the equality

$$n = l_1 + 2l_2 + \dots + pl_p + n',$$



**Corollary 5.17** *Square matrices  $A$  and  $B$  of order  $n$  are similar if and only if their eigenvalues coincide and for each eigenvalue  $\lambda$  and each  $i \leq n$ , we have*

$$\operatorname{rk}(A - \lambda E)^i = \operatorname{rk}(B - \lambda E)^i. \quad (5.29)$$

*Proof* The necessity of conditions (5.29) is obvious, since if  $A$  and  $B$  are similar, then so are the matrices  $(A - \lambda E)^i$  and  $(B - \lambda E)^i$ , which means that their ranks are the same.

We now prove sufficiency. Suppose that the conditions (5.29) are satisfied. We shall construct transformations  $\mathcal{A} : L \rightarrow L$  and  $\mathcal{B} : L \rightarrow L$  having in some basis  $e_1, \dots, e_n$  of the vector space  $L$  the matrices  $A$  and  $B$ . Let the transformation  $\mathcal{A}$  be brought into Jordan normal form in some basis  $f_1, \dots, f_n$ , and the same for  $\mathcal{B}$  in some basis  $g_1, \dots, g_n$ . In view of equality (5.29) and using formulas (5.25), we conclude that these Jordan forms coincide. This means that the matrices  $A$  and  $B$  are similar to some third matrix, and consequently, by transitivity, they are similar to each other.  $\square$

As an additional application of formulas (5.27), let us determine when a matrix can be brought into diagonal form, which is a special case of Jordan form in which all the Jordan blocks are of order 1. In other words, all the cyclic subspaces are of dimension one. This means that  $l_2 = \dots = l_n = 0$ . From the second equality in formulas (5.27), it follows that for this, it is necessary and sufficient that the condition  $r_1 = r_2$  be satisfied (for sufficiency, we must use the fact that  $l_i \geq 0$ ). We have thus proved the following criterion.

**Theorem 5.18** *A linear transformation  $\mathcal{A}$  can be brought into diagonal form if and only if for every one of its eigenvalues  $\lambda$ , we have*

$$\operatorname{rk}(\mathcal{A} - \lambda \mathcal{E}) = \operatorname{rk}(\mathcal{A} - \lambda \mathcal{E})^2.$$

Of course, an analogous criterion holds for matrices.

## 5.4 Real Vector Spaces

Up to this point, we have been considering linear transformations of *complex* vector spaces (this is related to the fact that we have continually relied on the existence of an eigenvector for every linear transformation, which may not be true in the real case). However, the theory that we have built up gives us a great deal of information about the case of transformations of real vector spaces as well, which are especially important in applications.

Let us assume that the real vector space  $L_0$  is embedded in the complex vector space  $L$ , for example its complexification (as was done in Sect. 4.3), while a linear transformation  $\mathcal{A}_0$  of the space  $L_0$  determines a real linear transformation  $\mathcal{A}$  of the space  $L$ . In this section and the following one, a bar will denote complex conjugation.

**Theorem 5.19** *In the decomposition of the space  $L$  into cyclic subspaces with respect to the real linear transformation  $\mathcal{A}$ , the number of cyclic  $m$ -dimensional subspaces associated with the eigenvalue  $\lambda$  is equal to the number of cyclic  $m$ -dimensional subspaces associated with the complex-conjugate eigenvalue  $\bar{\lambda}$ .*

*Proof* Since the characteristic polynomial of a real transformation  $\mathcal{A}$  has real coefficients, it follows that for each root  $\lambda$ , the number  $\bar{\lambda}$  is also a root of the characteristic polynomial. Let us denote, as we did in the proof of Theorem 5.15, the number of cyclic  $m$ -dimensional subspaces for the eigenvalue  $\lambda$  by  $l_m$ , and the number of cyclic  $m$ -dimensional subspaces for the eigenvalue  $\bar{\lambda}$  by  $l'_m$ . In addition, we define  $r_i = \text{rk}(\mathcal{A} - \lambda\mathcal{E})^i$  and  $r'_i = \text{rk}(\mathcal{A} - \bar{\lambda}\mathcal{E})^i$ . Formulas (5.28) express the numbers  $l_m$  in terms of  $r_m$ . Since these formulas hold for every eigenvalue, they also express the numbers  $l'_m$  in terms of  $r'_m$ . Consequently, it suffices to show that  $r'_i = r_i$ , from which it will follow that  $l'_i = l_i$ , which is the assertion of the theorem.

To this end, we consider some basis of the space  $L_0$  (as a real vector space). It will also be a basis of the space  $L$  (as a complex vector space). Let  $A$  be the matrix of the linear transformation  $\mathcal{A}$  in this basis. By definition, it coincides with the matrix of the linear transformation  $\mathcal{A}_0$  in the same basis, and therefore, it consists of real numbers. Hence the matrix  $A - \bar{\lambda}E$  is obtained from  $A - \lambda E$  by replacing all the elements by their complex conjugates. We shall write this as

$$A - \bar{\lambda}E = \overline{A - \lambda E}.$$

It is easy to see that from this, it follows that for every  $i > 0$ , the equation

$$(A - \bar{\lambda}E)^i = \overline{(A - \lambda E)^i}$$

is satisfied. Thus our assertion is reduced to the following: if  $B$  is a matrix with complex elements and the matrix  $\bar{B}$  is obtained from  $B$  by replacing all its elements with their complex conjugates, then  $\text{rk } B = \text{rk } \bar{B}$ . The proof of this follows at once, however, from the definition of the rank of a matrix as the maximal order of the nonzero minors: indeed, it is clear that the minors of the matrix  $\bar{B}$  are obtained by complex conjugation from the minors of  $B$  with the same indices of rows and columns, which completes the proof of the theorem.  $\square$

Thus according to Theorem 5.19, the Jordan normal form (5.19) of a real linear transformation consists of Jordan blocks (5.20) corresponding to real eigenvalues  $\lambda_i$  and pairs of Jordan blocks of the same order corresponding to complex-conjugate pairs of eigenvalues  $\lambda_i$  and  $\bar{\lambda}_i$ .

Let us see what this gives us for the classification of linear transformations of a real vector space  $L_0$ . Let us consider the simple example of the case  $\dim L_0 = 2$ . By Theorem 5.19, the Jordan normal form of the linear transformation  $\mathcal{A}$  of the complex space  $L$  can have one of the three following forms:

$$(a) \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}, \quad (b) \begin{pmatrix} \alpha & 0 \\ 1 & \alpha \end{pmatrix}, \quad (c) \begin{pmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{pmatrix},$$

where  $\alpha$  and  $\beta$  are real, and  $\lambda$  is a complex, not real, number, that is,  $\lambda = a + ib$ , where  $i^2 = -1$  and  $b \neq 0$ .

In cases (a) and (b), as can be seen from the definition of the linear transformation  $\mathcal{A}$ , the matrix of the transformation  $\mathcal{A}_0$  already has the indicated form in some basis of the real vector space  $L_0$ .

As we showed in Sect. 4.3, in case (c), the transformation  $\mathcal{A}_0$  has in some basis the matrix

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

Thus we see that an arbitrary linear transformation of a two-dimensional real vector space has in some basis one of three forms:

$$(a) \quad \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}, \quad (b) \quad \begin{pmatrix} \alpha & 0 \\ 1 & \alpha \end{pmatrix}, \quad (c) \quad \begin{pmatrix} a & -b \\ b & a \end{pmatrix}, \quad (5.30)$$

where  $\alpha, \beta, a, b$  are real numbers and  $b \neq 0$ . By formula (3.43), this implies that an arbitrary real square matrix of order 2 is similar to a matrix having one of the three forms of (5.30).

In a completely analogous way, we may study the general case of linear transformations in a real vector space of arbitrary dimension.<sup>1</sup> By the same line of argument, one can show that every real square matrix is similar to a block-diagonal matrix

$$A = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_r \end{pmatrix},$$

where  $A_i$  is either a Jordan block (5.20) with a real eigenvalue  $\lambda_i$  or a matrix of even order having the block form

$$A_i = \begin{pmatrix} \Lambda_i & 0 & 0 & \cdots & \cdots & 0 \\ E & \Lambda_i & 0 & \cdots & \cdots & 0 \\ 0 & E & \Lambda_i & & & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \Lambda_i & 0 \\ 0 & 0 & \cdots & \cdots & E & \Lambda_i \end{pmatrix},$$

---

<sup>1</sup>One may find a detailed proof in, for example, the book *Lectures on Algebra*, by D.K. Faddeev (in Russian) or in Sect. 3.4 of *Matrix Analysis*, by Roger Horn and Charles Johnson. See the references section for details.



in which the blocks  $A_i$  and  $E$  are matrices of order 2:

$$A_i = \begin{pmatrix} a_i & -b_i \\ b_i & a_i \end{pmatrix}, \quad E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

## 5.5 Applications\*

For a matrix  $A$  in Jordan normal form, it is easy to calculate the value of  $f(A)$ , where  $f(x)$  is any polynomial of degree  $n$ . First of all, let us note that if the matrix  $A$  is in block-diagonal form

$$A = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_r \end{pmatrix}$$

with arbitrary blocks  $A_1, \dots, A_r$ , then

$$f(A) = \begin{pmatrix} f(A_1) & 0 & \cdots & 0 \\ 0 & f(A_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f(A_r) \end{pmatrix}.$$

This follows immediately from the decomposition of the space  $L$  as  $L = L_1 \oplus \cdots \oplus L_r$ , a direct sum of invariant subspaces, and from the fact that a linear transformation with matrix  $A$  defines on  $L_i$  a linear transformation with matrix  $A_i$ .

Thus it remains only to consider the case that  $A$  is a Jordan block, that is,

$$A = \begin{pmatrix} \lambda & 0 & 0 & \cdots & \cdots & 0 \\ 1 & \lambda & 0 & & & 0 \\ 0 & 1 & \lambda & & & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \lambda & 0 \\ 0 & 0 & \cdots & \cdots & 1 & \lambda \end{pmatrix}. \quad (5.31)$$

It will be convenient to represent it in the form  $A = \lambda E + B$ , where

$$B = \begin{pmatrix} 0 & 0 & 0 & \cdots & \cdots & 0 \\ 1 & 0 & 0 & & & 0 \\ 0 & 1 & 0 & & & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & 0 & 0 \\ 0 & 0 & \cdots & \cdots & 1 & 0 \end{pmatrix}. \quad (5.32)$$

Let us now write down Taylor's formula for a polynomial of degree  $n$ :

$$f(x + y) = f(x) + f'(x)y + \frac{f''(x)}{2!}y^2 + \cdots + \frac{f^{(n)}(x)}{n!}y^n. \quad (5.33)$$

We note that for the derivation of formula (5.33), we have to compute the binomial expansion of  $(x + y)^k$ ,  $k = 2, \dots, n$ , and then, of course, use commutativity of multiplication of numbers. If the commutative property did not hold, then we would not be able to obtain, for example, the expression  $(x + y)^2 = y^2 + 2xy + x^2$ , but only  $(x + y)^2 = y^2 + yx + xy + x^2$ . Therefore, in formula (5.33), we may replace  $x$  and  $y$  by numbers, but not by arbitrary matrices, instead only those that commute.

Let us substitute in formula (5.33) the arguments  $x = \lambda E$  and  $y = B$ , since the matrices  $\lambda E$  and  $B$  obviously commute. As is easily verified, for an arbitrary polynomial  $f(\lambda E) = f(\lambda)E$ , we obtain the expression

$$f(A) = f(\lambda)E + f'(\lambda)B + \frac{f''(\lambda)}{2!}B^2 + \cdots + \frac{f^{(n)}(\lambda)}{n!}B^n. \quad (5.34)$$

We now observe that in the basis  $e_1, \dots, e_m$  of the cyclic subspace generated by the principal vector  $e$  of grade  $m$ , the transformation  $\mathcal{B}$  with  $B$  of the form (5.32) assumes the following form:

$$\mathcal{B}(e_i) = \begin{cases} e_{i+1} & \text{for } i \leq m-1, \\ \mathbf{0} & \text{for } i > m-1. \end{cases}$$

Applying the formula  $k$  times, we obtain that

$$\mathcal{B}^k(e_i) = \begin{cases} e_{i+k} & \text{for } i \leq m-k, \\ \mathbf{0} & \text{for } i > m-k. \end{cases}$$

From this, it is clear that the matrix  $B^k$  has the following very simple form:

$$B^k = \begin{pmatrix} 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & & & & & & \vdots \\ 1 & 0 & & & & & & \vdots \\ 0 & 1 & & & & & & \vdots \\ 0 & 0 & \ddots & & & & & \vdots \\ \vdots & \vdots & & \ddots & & & & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \end{pmatrix}.$$

In order to describe this in words, we shall call the collection of elements  $a_{ij}$  in the matrix  $A = (a_{ij})$  with  $i = j$  the *main diagonal*, while the collection of elements  $a_{ij}$  with  $i - j = k$  (where  $k$  is a given number) forming a diagonal parallel to the main diagonal will be called the *diagonal lying  $k$  steps from the main diagonal*. Thus in the matrix  $B^k$ , the diagonal lying  $k$  steps from the main diagonal contains all 1's, while the remaining matrix entries are zero.

Formula (5.34) now gives for a Jordan block  $A$  of order  $m$  the expression

$$f(A) = \begin{pmatrix} \varphi_0 & 0 & 0 & \cdots & 0 & 0 \\ \varphi_1 & \varphi_0 & 0 & \cdots & 0 & 0 \\ \varphi_2 & \varphi_1 & \varphi_0 & \ddots & & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \varphi_{m-2} & \varphi_{m-3} & \ddots & \ddots & \varphi_0 & 0 \\ \varphi_{m-1} & \varphi_{m-2} & \varphi_{m-3} & \cdots & \varphi_1 & \varphi_0 \end{pmatrix}, \quad (5.35)$$

where  $\varphi_k = f^{(k)}(\lambda)/k!$ , that is, the numbers  $\varphi_k$  are the coefficients in the Taylor expansion (5.34).

Let us look at a very simple example. Suppose we wish to raise a matrix  $A$  of order 2 to a very high power  $p$  (for example,  $p = 2000$ ). To perform such calculations by hand seems hopeless. But the theory that we have constructed proves here to be very useful. Let us find an eigenvalue of the linear transformation  $\mathcal{A}$  with matrix  $A$ , that is, a root of the second-degree trinomial  $|A - \lambda E|$ . Here two cases are possible.

*Case 1.* The trinomial  $|A - \lambda E|$  has distinct roots  $\lambda_1$  and  $\lambda_2$ . We can easily find the associated eigenvectors  $\mathbf{e}_1$  and  $\mathbf{e}_2$ , for which

$$(\mathcal{A} - \lambda_1 \mathcal{E})(\mathbf{e}_1) = \mathbf{0}, \quad (\mathcal{A} - \lambda_2 \mathcal{E})(\mathbf{e}_2) = \mathbf{0}.$$

As we know, the vectors  $\mathbf{e}_1$  and  $\mathbf{e}_2$  are linearly independent, and in the basis  $\mathbf{e}_1, \mathbf{e}_2$ , the transformation  $\mathcal{A}$  has the diagonal matrix  $\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ . If  $C$  is the transition matrix

from the original basis in which the transformation  $\mathcal{A}$  has matrix  $A$  to the basis  $\mathbf{e}_1, \mathbf{e}_2$ , then

$$A = C^{-1} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} C, \quad (5.36)$$

whence is easily obtained for any  $p$  (as large as desired), the formula

$$A^p = C^{-1} \begin{pmatrix} \lambda_1^p & 0 \\ 0 & \lambda_2^p \end{pmatrix} C. \quad (5.37)$$

Let us now consider the second case.

*Case 2.* The trinomial  $|A - \lambda E|$  has a multiple root  $\lambda$  (which therefore must be real). Then the Jordan normal form of the matrix  $A$  has the form of a single block  $\begin{pmatrix} \lambda & 0 \\ 1 & \lambda \end{pmatrix}$  or  $\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$ . In the latter variant, the Jordan normal form of the matrix is equal to  $\lambda E$ , and therefore the matrix  $A$  is also equal to  $\lambda E$  (this follows, for example, from the fact that if in some basis, a linear transformation has the matrix  $\lambda E$ , then it will have the same matrix in every other basis as well). Thus in this last variant we are dealing with the previous case, in which  $\lambda_1 = \lambda_2 = \lambda$ , and the calculation of  $A^p$  is obtained by formula (5.37), where we have only to substitute  $\lambda_1$  and  $\lambda_2$  for  $\lambda$ . It remains to consider the first variant. For a Jordan block  $\begin{pmatrix} \lambda & 0 \\ 1 & \lambda \end{pmatrix}$ , by formula (5.35), we obtain

$$\begin{pmatrix} \lambda & 0 \\ 1 & \lambda \end{pmatrix}^p = \begin{pmatrix} \lambda^p & 0 \\ p\lambda^{p-1} & \lambda^p \end{pmatrix}.$$

If  $\mathbf{e}_1, \mathbf{e}_2$  are vectors such that

$$(\mathcal{A} - \lambda \mathcal{E})(\mathbf{e}_1) \neq \mathbf{0}, \quad \mathbf{e}_2 = (\mathcal{A} - \lambda \mathcal{E})(\mathbf{e}_1),$$

then in the basis  $\mathbf{e}_1, \mathbf{e}_2$ , the matrix of the transformation  $\mathcal{A}$  is in Jordan normal form. We denote by  $C$  the transition matrix to this basis, and using the transition formula

$$A = C^{-1} \begin{pmatrix} \lambda & 0 \\ 1 & \lambda \end{pmatrix} C,$$

we obtain

$$A^p = C^{-1} \begin{pmatrix} \lambda^p & 0 \\ p\lambda^{p-1} & \lambda^p \end{pmatrix} C. \quad (5.38)$$

Formulas (5.37) and (5.38) solve our problem.

We can now apply the same ideas not only to polynomials, but to other functions, for example those given by a convergent power series. Such functions are called *analytic*. To do this, we need the concept of *convergence* of a sequence of matrices. Let us recall that the notion of convergence for a sequence of square matrices of a given order with real coefficients was defined earlier, in Sect. 4.4. Moreover, in that same section, we introduced on the set of such matrices the metric  $r(A, B)$ , after converting it to a metric space, on which the notion of convergence is defined

automatically (see p. xvii). It is obvious that the metric  $r(A, B)$  defined by formulas (4.36) and (4.37) is also a metric on the set of square matrices of a given order with complex coefficients, and therefore transforms it into a metric space.

With this definition, the convergence of a sequence of matrices  $A^{(k)} = (a_{ij}^{(k)})$ ,  $k = 1, 2, \dots$ , to a matrix  $B = (b_{ij})$  means that  $a_{ij}^{(k)} \rightarrow b_{ij}$  for  $k \rightarrow \infty$  for all  $i, j$ . In this case, we write  $A^{(k)} \rightarrow B$  for  $k \rightarrow \infty$  or  $\lim_{k \rightarrow \infty} A^{(k)} = B$ . The matrix  $B$  is called the *limit* of the sequence  $A^{(k)}$ ,  $k = 1, 2, \dots$ . Similarly, we can define the limit of a family of matrices  $A(h)$  depending on a parameter  $h$  assuming values that are not necessarily natural numbers (as was the case for a sequence), but real values, and approaching an arbitrary value  $h_0$ . By definition,  $\lim_{h \rightarrow h_0} A(h) = B$  if  $\lim_{h \rightarrow h_0} r(A(h), B) = 0$ . In other words, this means that  $\lim_{h \rightarrow h_0} a_{ij}(h) = b_{ij}$  for all  $i, j$ .

Just as in the case of numbers, once we have the notion of convergence of a sequence of matrices, it is possible to talk about the convergence of *series* of matrices. Without any alteration, we can transfer theorems on series known from analysis to series of matrices. Let the function  $f(x)$  be defined by the power series

$$f(x) = \alpha_0 + \alpha_1 x + \dots + \alpha_k x^k + \dots \quad (5.39)$$

Then by definition,

$$f(A) = \alpha_0 E + \alpha_1 A + \dots + \alpha_k A^k + \dots \quad (5.40)$$

Suppose the power series (5.39) converges for  $|x| < r$  and the matrix  $A$  is in the form of a Jordan block (5.31) with eigenvalue  $\lambda$ , of absolute value less than  $r$ . Then, examining the sum of the first  $k$  terms of the series (5.40) and passing to the limit  $k \rightarrow \infty$ , we obtain that the series (5.40) converges, and for  $f(A)$ , formula (5.35) holds. If we now take a matrix  $A'$  similar to some Jordan block  $A$ , that is, related to it by  $A' = C^{-1}AC$ , where  $C$  is some nonsingular matrix, then from the obvious relationship  $(C^{-1}AC)^k = C^{-1}A^kC$ , we obtain from (5.40) that

$$f(A') = C^{-1}(\alpha_0 E + \alpha_1 A + \dots + \alpha_k A^k + \dots)C = C^{-1}f(A)C. \quad (5.41)$$

Formulas (5.35) and (5.41) allow us to compute  $f(A)$  for any analytic function  $f(x)$ . Using results from analysis, we can extend the notion of functions of matrices to a wider class of functions (for example, to continuous functions with the help of the theorem on uniform approximation of continuous functions by polynomials). However, we shall not address these questions here.

In applications, of especial importance are *exponentials* of matrices. We recall that the exponential function of a number  $x$  can be defined by the series summation

$$e^x = 1 + x + \frac{1}{2!}x^2 + \dots + \frac{1}{k!}x^k + \dots, \quad (5.42)$$

which, as proved in a course in analysis, converges for all real or complex numbers  $x$ . According to this, the *exponential* of a matrix  $A$  is defined by the series

$$e^A = E + A + \frac{1}{2!}A^2 + \cdots + \frac{1}{k!}A^k + \cdots, \quad (5.43)$$

which converges for every matrix  $A$  with real or complex entries.

Let us verify that if matrices  $A$  and  $B$  commute, then a basic property of the numerical exponential function is transferred to the matrix exponential function:

$$e^A e^B = e^{A+B}. \quad (5.44)$$

Indeed, substituting into the left-hand side of (5.44) the expressions (5.43) for  $e^A$  and  $e^B$ , removing parentheses, and collecting like terms, we obtain

$$\begin{aligned} e^A e^B &= \left( E + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots \right) \left( E + B + \frac{1}{2!}B^2 + \frac{1}{3!}B^3 + \cdots \right) \\ &= E + (A + B) + \left( \frac{1}{2!}A^2 + AB + \frac{1}{2!}B^2 \right) \\ &\quad + \left( \frac{1}{3!}A^3 + \frac{1}{2!}A^2B + \frac{1}{2!}AB^2 + \frac{1}{3!}B^3 \right) + \cdots \\ &= E + (A + B) + \frac{1}{2!}(A + B)^2 + \frac{1}{3!}(A + B)^3 + \cdots, \end{aligned}$$

which coincides with the expression (5.43) for  $e^{A+B}$ . As justification for the generalization made above, it is necessary to note that first of all, as is known from analysis, for the corresponding exponential function (5.43), the numeric series (5.42) converges absolutely on the entire real axis (this allows the terms to be summed in arbitrary order), and second, matrices  $A$  and  $B$  commute (without this, this last generalization would be impossible, which we know by virtue of what we discussed earlier on page 177).

In particular, from (5.44) follows the important relationship

$$e^{A(t+s)} = e^{At} e^{As} \quad (5.45)$$

for all numbers  $t$  and  $s$  and every square matrix  $A$ . From this, it is easy to derive that

$$\frac{d}{dt} e^{At} = A e^{At} \quad (5.46)$$

(understanding that differentiation of the matrix function is to be taken element-wise).

Indeed, by the definition of differentiation,

$$\frac{d}{dt} e^{At} = \lim_{h \rightarrow 0} \frac{e^{A(t+h)} - e^{At}}{h},$$

while from (5.45), it follows that

$$\frac{e^{A(t+h)} - e^{At}}{h} = \frac{e^{Ah} e^{At} - e^{At}}{h} = \frac{e^{Ah} - E}{h} e^{At}.$$

Finally, from (5.43) we easily obtain the equality

$$\lim_{h \rightarrow 0} \frac{e^{Ah} - E}{h} = \lim_{h \rightarrow 0} h^{-1} \left( (Ah) + \frac{1}{2!} (Ah)^2 + \cdots + \frac{1}{k!} (Ah)^k + \cdots \right) = A.$$

All these considerations have numerous applications in the theory of differential equations. Let us consider a system of  $n$  linear homogeneous differential equations

$$\frac{dx_i}{dt} = \sum_{j=1}^n a_{ij} x_j, \quad i = 1, \dots, n, \quad (5.47)$$

where  $a_{ij}$  are certain constant coefficients and  $x_i = x_i(t)$  are unknown differentiable functions of the variable  $t$ . Similarly to what was done earlier for systems of linear algebraic equations (Example 2.49, p. 62), the system of linear differential equations (5.47) can also be written down compactly in matrix form if we introduce the column vectors

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \frac{d\mathbf{x}}{dt} = \begin{pmatrix} dx_1/dt \\ \vdots \\ dx_n/dt \end{pmatrix}$$

and a square matrix of order  $n$  consisting of the coefficients of the system:  $A = (a_{ij})$ . Then system (5.47) can be written in the form

$$\frac{d\mathbf{x}}{dt} = A\mathbf{x}. \quad (5.48)$$

The number  $n$  is called the *order* of this system.

For any constant vector  $\mathbf{x}_0$ , let us consider the vector  $\mathbf{x}(t) = e^{At} \mathbf{x}_0$ , depending on the variable  $t$ . This vector satisfies the system (5.48). Indeed, for arbitrary matrices  $A(t)$  and  $B$  (possibly rectangular, provided that the number of columns of  $A(t)$  coincides with the number of rows of  $B$ ), if only the matrix  $B$  is constant, one has the equality

$$\frac{d}{dt}(A(t)B) = \frac{dA(t)}{dt}B,$$

after which it remains to use relationship (5.46). Similarly, for arbitrary matrices  $A(t)$  and  $B$ , where  $B$  is constant and the number of columns of  $B$  coincides with the number of rows of  $A(t)$ , we have the formula

$$\frac{d}{dt}(BA(t)) = B \frac{dA(t)}{dt}. \quad (5.49)$$

Since with  $t = 0$ , the matrix  $e^{At}$  equals  $E$ , the solution  $\mathbf{x}(t) = e^{At}\mathbf{x}_0$  satisfies the initial condition  $\mathbf{x}(0) = \mathbf{x}_0$ . But the uniqueness theorem proved in the theory of differential equations asserts that for a given  $\mathbf{x}_0$ , such a solution is unique. Thus we may obtain *all* solutions of the system (5.48) in the form  $e^{At}\mathbf{x}_0$  if we consider the vector  $\mathbf{x}_0$  not as fixed, but as taking all possible values in a space of dimension  $n$ .

Finally, it is also possible to obtain an explicit formula for the solutions. To this end, let us make a linear substitution of variables in the system of equations (5.48) according to the formula  $\mathbf{y} = C^{-1}\mathbf{x}$ , where  $C$  is a nonsingular constant square matrix of order  $n$ . Then taking into account relationships (5.49), (5.48), and  $\mathbf{x} = C\mathbf{y}$ , we obtain

$$\frac{d\mathbf{y}}{dt} = C^{-1} \frac{d\mathbf{x}}{dt} = C^{-1} A \mathbf{x} = (C^{-1} A C) \mathbf{y}. \quad (5.50)$$

Formula (5.50) shows that the matrix  $A$  of a system of linear differential equations under a linear replacement of variables changes according to the same law as the matrix of a linear transformation under a suitable change of basis. In accord with what we have done in previous sections, we may choose as  $C$  a matrix with whose help, the matrix  $A$  is converted to Jordan normal form. As a result, the system (5.48) can be rewritten in the form

$$\frac{d\mathbf{y}}{dt} = A' \mathbf{y}, \quad (5.51)$$

where the matrix  $A' = C^{-1} A C$  is in Jordan normal form.

Let

$$A' = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_r \end{pmatrix}, \quad (5.52)$$

where the  $A_i$  are Jordan blocks. Then system (5.51) is decomposed into  $r$  systems

$$\frac{d\mathbf{y}_i}{dt} = A_i \mathbf{y}_i, \quad i = 1, \dots, r,$$

and for each of these, we can express the solution in the form  $e^{A_i t} \mathbf{x}_0^{(i)}$  and find the matrix  $e^{A_i t}$  from the relationship (5.35). Here  $f(x) = e^{xt}$ , and consequently,

$$f^{(k)}(x) = \frac{d^k}{dx^k} e^{xt} = t^k e^{xt}, \quad \varphi_k = \frac{t^k}{k!} e^{\lambda t}.$$



This implies that for blocks  $A_i$  of the form (5.31) of order  $m$ , formula (5.35) gives us

$$e^{At} = e^{\lambda t} \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ t & 1 & 0 & \cdots & 0 & 0 \\ \frac{t^2}{2} & t & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \frac{t^{m-2}}{(m-2)!} & \frac{t^{m-3}}{(m-3)!} & \ddots & \ddots & 1 & 0 \\ \frac{t^{m-1}}{(m-1)!} & \frac{t^{m-2}}{(m-2)!} & \frac{t^{m-3}}{(m-3)!} & \cdots & t & 1 \end{pmatrix}. \quad (5.53)$$

This implies that the solutions of the system (5.48) can be decomposed into series whose lengths are equal to the orders of the Jordan blocks in the representation (5.52), and for a block of order  $m$ , all solutions of the given series can be expressed as linear combinations (with constant coefficients) of the functions

$$e^{\lambda t}, \quad t e^{\lambda t}, \quad \dots, \quad t^{m-1} e^{\lambda t}. \quad (5.54)$$

It is easily verified that the collection of solutions of system (5.48) forms a vector space, where the addition of two vectors and multiplication of a vector by a scalar are defined just as were addition and multiplication by a scalar of the corresponding functions. The set of functions (5.54) forms a basis of the space of solutions of the system (5.48). In the theory of differential equations, such a set is called a *fundamental system of solutions*.

In conclusion, let us say a few words about linear differential equations with real coefficients in the plane ( $n = 2$ ) (that is, assuming that in system (5.48), the matrix  $A$  and vector  $\mathbf{x}$  are real). Here, we should distinguish four possibilities for the matrix  $A$  and roots of the polynomial  $|A - \lambda E|$ :

- (a) The roots are real and distinct: ( $\alpha$  and  $\beta$ ).
- (b) There is a multiple root  $\alpha$  (necessarily real) and  $A = \alpha E$ .
- (c) There is a multiple root  $\alpha$ , but  $A \neq \alpha E$ .
- (d) The roots are complex conjugate:  $a + ib$  and  $a - ib$  (here  $i^2 = -1$  and  $b \neq 0$ ).

In each of these cases, the matrix  $A$  can be brought (by multiplication on the left by  $C^{-1}$  and on the right by  $C$ , where  $C$  is some nonsingular real matrix) into the following normal forms:

$$(a) \quad \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}, \quad (b) \quad \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix}, \quad (c) \quad \begin{pmatrix} \alpha & 0 \\ 1 & \alpha \end{pmatrix}, \quad (d) \quad \begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

The solution  $\mathbf{x}(t)$  of the associated differential equation is obtained in the form  $\mathbf{x}(t) = e^{At} \mathbf{x}_0$ , where  $\mathbf{x}_0 = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$  is the vector of the original data. Further, we can use formula (5.53), considering that the matrix  $A$  of the system has the normal form (a), (b), (c), or (d). Here in cases (a)–(c), we will obtain

$$(a) \quad \mathbf{x}(t) = \begin{pmatrix} e^{\alpha t} c_1 \\ e^{\beta t} c_2 \end{pmatrix}, \quad (b) \quad \mathbf{x}(t) = \begin{pmatrix} e^{\alpha t} c_1 \\ e^{\alpha t} c_2 \end{pmatrix}, \quad (5.55)$$

$$(c) \quad \mathbf{x}(t) = \begin{pmatrix} e^{\alpha t} & 0 \\ t e^{\alpha t} & e^{\alpha t} \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} c_1 e^{\alpha t} \\ c_1 t e^{\alpha t} + c_2 e^{\alpha t} \end{pmatrix}. \quad (5.56)$$

In case (d), we obtain  $\mathbf{x}(t) = e^{At} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ , where  $A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ . In Example 4.2 (p. 134) we established that  $A$  is the matrix of a linear transformation of the plane  $\mathbb{C}$  with complex variable  $z$  that multiplies  $z$  by the complex number  $a + ib$ . This means, by the definition of the exponential function, that  $e^{At}$  is the matrix of multiplication of  $z$  by the complex number  $e^{(a+ib)t}$ . By Euler's formula,

$$e^{(a+ib)t} = e^{at} (\cos bt + i \sin bt) = p + iq,$$

where  $p = e^{at} \cos bt$  and  $q = e^{at} \sin bt$ . Thus we obtain a linear transformation of the real plane  $\mathbb{C}$  with complex variable  $z$  that multiplies each complex number  $z \in \mathbb{C}$  by the given complex number  $p + iq$ . As we saw in Example 4.2, the matrix of such a linear transformation has the form (4.2). Multiplying it by the column vector  $\mathbf{x}_0$  of the original data and substituting the expressions  $p = e^{at} \cos bt$  and  $q = e^{at} \sin bt$ , we obtain our final formula:

$$(?) \quad \mathbf{x}(t) = \begin{pmatrix} p & -q \\ q & p \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = e^{at} \begin{pmatrix} c_1 \cos bt - c_2 \sin bt \\ c_1 \sin bt + c_2 \cos bt \end{pmatrix}. \quad (5.57)$$

The plane of variables  $(x_1, x_2)$  is called the *phase plane* of the system (5.48) for  $n = 2$ . Formulas (5.55)–(5.57) define (in parametric form) certain curves in the phase plane, where to each pair of values  $c_1, c_2$  there corresponds in general a curve passing through the point  $(c_1, c_2)$  of the phase plane for  $t = 0$ . These oriented curves (the orientation is given by the direction of motion corresponding to an increase in the parameter  $t$ ) are called *phase curves* of system (5.48), and the collection of all phase curves corresponding to all possible values of  $c_1, c_2$  is called the *phase portrait* of the system. Let us pose the following question: What does the phase portrait of the system (5.48) look like in cases (a)–(d)?

First of all, we note that among all solutions  $\mathbf{x}(t)$  there is always the constant  $\mathbf{x}(t) \equiv \mathbf{0}$ . It is obtained by substituting in formulas (5.55)–(5.57) the initial values  $c_1 = c_2 = 0$ . The phase curve corresponding to this solution is simply the point  $x_1 = x_2 = 0$ . Constant solutions (and their corresponding phase curves, points in the phase plane) are called *singular points* or *equilibrium points* or *fixed points* of the differential equation.<sup>2</sup> Similarly, just as the study of a function usually begins with a search for its extreme points, so a study of a differential equation usually begins with a search for its singular points.

Are there singular points of system (5.48) other than  $x_1 = x_2 = 0$ ? Singular points are the constant solutions of a system of equations, and since the derivative of a constant solution is identically equal to zero (that is, the left-hand side of system (5.48) is identically zero), this means that the right-hand side of system (5.48) must also be identically equal to zero. Therefore, singular points are precisely the

<sup>2</sup>This name comes from the fact that if at some moment in time, a material point whose motion is described by system (5.48) is located at a singular point, then it will remain there forever.

solutions of the system of linear homogeneous equations  $A\mathbf{x} = \mathbf{0}$ . If the matrix  $A$  is nonsingular, then the system  $A\mathbf{x} = \mathbf{0}$  has no solutions other than the null solution, and therefore, system (5.48) has no singular points other than  $x_1 = x_2 = 0$ . If the matrix  $A$  is singular and its rank is equal to 1, then system (5.48) has an infinite number of singular points lying on a line in the phase plane. But in the case that the rank of the matrix  $A$  is equal to 0, all points of the phase plane are singular points.

In the sequel, we will consider that the matrix  $A$  is nonsingular and examine what sorts of phase portraits they correspond to in the cases (a)–(d) presented above. In all the figures, the  $x$ -axis corresponds to the variable  $x_1$ , while the  $y$ -axis represents the variable  $x_2$ .

(a) The roots  $\alpha$  and  $\beta$  are real and distinct. In this case, there are three possibilities:  $\alpha$  and  $\beta$  have different signs, both are negative, or both are positive.

(a.1) If  $\alpha$  and  $\beta$  have different signs, then a singular point is called a *saddle*. For definiteness, let us assume that  $\alpha < 0$  and  $\beta > 0$ . To the initial value  $c_1 \neq 0$ ,  $c_2 = 0$  there corresponds the solution  $x_1(t) = c_1 e^{\alpha t}$ ,  $x_2(t) = 0$ , passing through the point  $(c_1, 0)$  at  $t = 0$ . The associated phase curve is the horizontal ray  $x_1 > 0$ ,  $x_2 = 0$  (if  $c_1 > 0$ ) or  $x_1 < 0$ ,  $x_2 = 0$  (if  $c_1 < 0$ ) such that the direction along the curve with increasing  $t$  is toward the singular point  $x_1 = x_2 = 0$ .

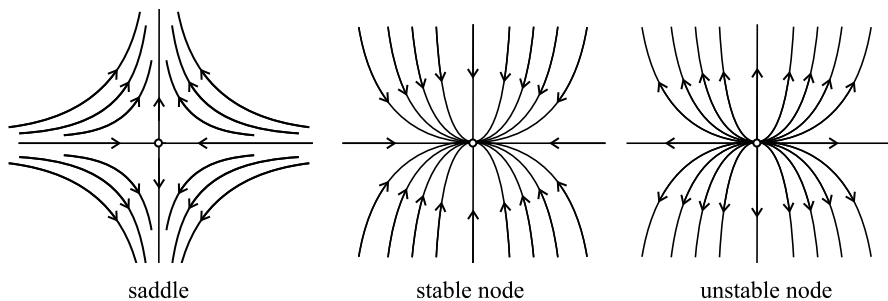
Similarly, to the initial point  $c_1 = 0$ ,  $c_2 \neq 0$  corresponds the solution  $x_1(t) = 0$ ,  $x_2(t) = c_2 e^{\beta t}$ , passing through the point  $(0, c_2)$  at  $t = 0$ . The associated phase curve is the vertical ray  $x_1 = 0$ ,  $x_2 > 0$  (if  $c_2 > 0$ ) or  $x_1 = 0$ ,  $x_2 < 0$  (if  $c_2 < 0$ ) such that the direction along the curve for increasing  $t$  is away from the singular point  $x_1 = x_2 = 0$ .

Thus there are two phase curves asymptotically approaching the singular point as  $t \rightarrow +\infty$  (they are called *stable separatrices*), and two curves approaching it for  $t \rightarrow -\infty$  (they are called *unstable separatrices*). Let us make one crucial observation: from the fact that  $e^{\alpha t} \rightarrow 0$  for  $t \rightarrow +\infty$  and  $e^{\beta t} \rightarrow 0$  for  $t \rightarrow -\infty$ , it follows that stable and unstable separatrices approach a saddle arbitrarily closely as  $t \rightarrow +\infty$  and  $t \rightarrow -\infty$  respectively but never reach it in *finite* time.

The stable and unstable separatrices of a saddle partition the phase plane into four sectors. In our case (in which the matrix of system (5.48) is in Jordan form), the separatrices lie on the coordinate axes, and therefore, these sectors coincide with the Cartesian quadrants. Let us see how the remaining phase curves behave with respect to the initial values  $c_1 \neq 0$ ,  $c_2 \neq 0$ . We observe first that if the initial point  $(c_1, c_2)$  lies in any of the four sectors, then after passing through it for  $t = 0$ , the phase curve remains in that sector for all values of  $t$ . This follows obviously from the fact that the functions  $x_1(t) = c_1 e^{\alpha t}$  and  $x_2(t) = c_2 e^{\beta t}$  are of fixed sign.

For definiteness, let us consider the first quadrant  $c_1 > 0$ ,  $c_2 > 0$  (the other cases can be obtained from this one by a symmetry transformation with respect to the  $x$ - or  $y$ -axis or with respect to the origin). Let us raise the function  $x_1(t) = c_1 e^{\alpha t}$  to the  $\beta$  power, and the function  $x_2(t) = c_2 e^{\beta t}$  to the  $\alpha$  power. After dividing one by the other and canceling the factor  $e^{\alpha\beta t}$ , we obtain the relationship

$$\frac{x_1^\beta}{x_2^\alpha} = \frac{c_1^\beta}{c_2^\alpha} = c, \quad (5.58)$$



**Fig. 5.1** Saddle and nodes

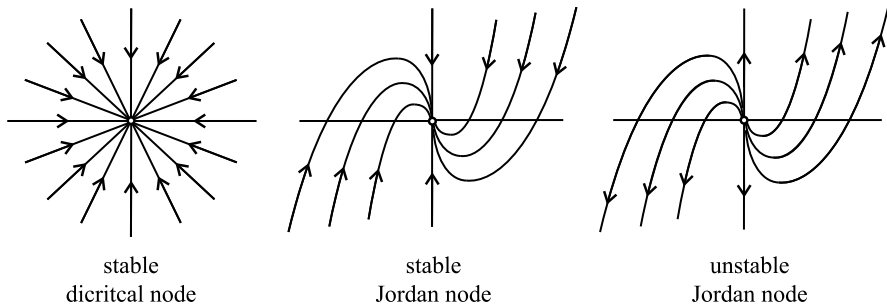
where the constant  $c$  is determined by the initial values  $c_1, c_2$ . Since the numbers  $\alpha$  and  $\beta$  have opposite signs, the phase curve in the plane  $(x_1, x_2)$  corresponding to this equation has a form similar to a hyperbola. This phase curve passes at some positive distance from the singular point  $x_1 = x_2 = 0$ , asymptotically approaching one of the unstable separatrices as  $t \rightarrow +\infty$  and to one of the stable separatrices as  $t \rightarrow -\infty$ . Such phase curves are said to be of *hyperbolic* or *saddle* type.

Thus in the case of a saddle, we have two stable separatrices approaching the singular point as  $t \rightarrow +\infty$  and two unstable separatrices approaching it as  $t \rightarrow -\infty$ , and also an infinite number of saddle-type phase curves filling the four sectors into which the separatrices divide the phase plane. The associated phase portrait is shown in Fig. 5.1.

(a.2) If  $\alpha$  and  $\beta$  have the same sign, then a singular point is called a *node*. Moreover, if  $\alpha$  and  $\beta$  are negative, then the node is said to be *stable*, while if  $\alpha$  and  $\beta$  are positive, the node is *unstable*. The reason for this terminology will soon become clear.

For definiteness, we will restrict our examination to stable nodes (unstable nodes are studied similarly), that is, we shall assume that the numbers  $\alpha$  and  $\beta$  are negative. As in the case of a saddle, the phase curve corresponding to the initial value  $c_1 \neq 0, c_2 = 0$  is the horizontal ray  $x_1 > 0, x_2 = 0$  (if  $c_1 > 0$ ) or  $x_1 < 0, x_2 = 0$  (if  $c_1 < 0$ ) such that the direction along the curve for increasing  $t$  is toward the singular point. The phase curve corresponding to the initial value  $c_1 = 0, c_2 \neq 0$  is the vertical ray  $x_1 = 0, x_2 > 0$  (if  $c_2 > 0$ ) or  $x_1 = 0, x_2 < 0$  (if  $c_2 < 0$ ) such that the direction along the curve for increasing  $t$  is also toward the singular point.

As in the case of a saddle, it is clear that if the initial point  $(c_1, c_2)$  lies in one of the four quadrants, then the phase curve passing through it for  $t = 0$  remains in that quadrant for all values of  $t$ . Let us consider the first quadrant  $c_1 > 0, c_2 > 0$ . Proceeding as we did in the case of a saddle, we again obtain the equation (5.58). But now the numbers  $\alpha$  and  $\beta$  have the same sign, and the phase curve corresponding to this equation has quite a different form from that in the case of a saddle. After a transformation of (5.58), we obtain the exponential function  $x_1 = c^{1/\beta} x_2^{\alpha/\beta}$ . If  $\alpha > \beta$ , then the exponent  $\alpha/\beta$  is greater than 1, and the graph of this function is similar to a branch of the parabola  $x_1 = x_2^2$ . However, if  $\alpha < \beta$ , then the exponent  $\alpha/\beta$  is less than 1, and the graph of the function looks like a branch of the parabola



**Fig. 5.2** Dicritical and Jordan nodes

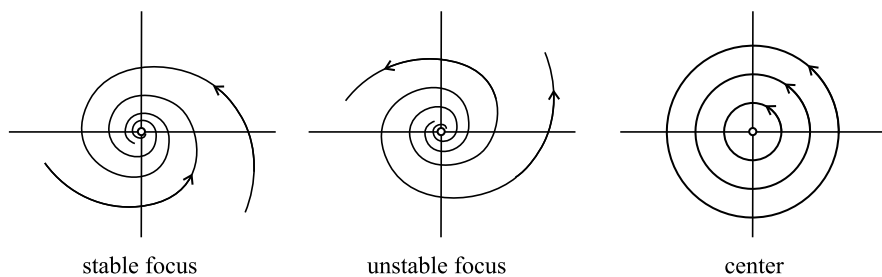
$x_2 = x_1^2$ . Thus in the case of a stable node, all the phase curves approach the singular point as  $t \rightarrow +\infty$ , while for  $t \rightarrow -\infty$ , they move away from it (for an unstable node we must exchange the positions of  $+\infty$  and  $-\infty$ ). Such phase curves are called *parabolic*. Phase portraits of stable and unstable nodes are depicted in Fig. 5.1.

It is now possible to explain the terminology *stable* and *unstable*. If a material point was located at an equilibrium point that was a stable node and was brought out from that point by some external action, then moving along the curve depicted in the phase portrait, it will strive to return to that position. But if it was an unstable node, then a material point brought out from an equilibrium point not only would not strive to return to that position, but on the contrary, it would move away from it with exponentially increasing speed.

(b) If a matrix  $A$  is similar to the matrix  $\alpha E$ , then a singular point is called a *dicritical node* or *bicritical node*. Proceeding in the same way as before, we obtain the relationship (5.58) with  $\beta = \alpha$ , from which follows the equation  $x_1/x_2 = c_1/c_2$ . All the phase curves are rays with origin at  $x_1 = x_2 = 0$ . Moreover, if  $\alpha < 0$ , then motion along them as  $t \rightarrow +\infty$  proceeds toward the equilibrium point  $x_1 = x_2 = 0$ , while if  $\alpha > 0$ , then away from it. Thus in the case  $\alpha < 0$  ( $\alpha > 0$ ), we have a stable (unstable) dicritical node. The phase portrait of a stable dicritical node is depicted in Fig. 5.2. In the case of an unstable dicritical node, it is necessary only to change the directions of the arrows to their opposite.

(c) If the solution to the equation is given by formula (5.56), then a singular point is called a *Jordan node*. If  $\alpha < 0$ , then the Jordan node is stable, and if  $\alpha > 0$ , then it is unstable. For  $c_1 \neq 0$ ,  $c_2 = 0$ , we obtain two phase curves, namely the horizontal rays  $x_1 > 0$ ,  $x_2 = 0$  and  $x_1 < 0$ ,  $x_2 = 0$ , whose motion is in the direction of the singular point for  $\alpha < 0$  and away from the singular point for  $\alpha > 0$ . In the investigation of phase curves for  $c_2 \neq 0$ , one must study the properties of the functions  $x_1(t) = c_1 e^{\alpha t}$  and  $x_2(t) = (c_1 t + c_2) e^{\alpha t}$  for  $c_1 > 0$  and for  $c_1 < 0$ . As a result, for a stable (unstable) Jordan node, one obtains the phase portrait depicted in Fig. 5.2. All the phase curves (except the two vertical rays) look like pieces of a parabola, each of which lies entirely either in the right or left half-plane and intersects the  $x$ -axis in a single point.

(d) The roots are complex conjugates:  $a + ib$  and  $a - ib$ , where  $b \neq 0$ . Here it is necessary to consider two cases:  $a \neq 0$  and  $a = 0$ .



**Fig. 5.3** Foci and center

(d.1) If  $a \neq 0$ , then a singular point is called a *focus*. In order to visualize the behavior of phase curves given by formula (5.57), we observe that the vector  $\mathbf{x}(t)$  is obtained from the vector  $\mathbf{x}_0$  with coordinates  $(c_1, c_2)$  by rotating it through the angle  $bt$  and multiplying by  $e^{at}$ . Therefore, the phase curves are spirals that “wind” around the singular point  $x_1 = x_2 = 0$  as  $t \rightarrow +\infty$  (if  $a < 0$ ) or as  $t \rightarrow -\infty$  (if  $a > 0$ ). For  $a < 0$  and  $a > 0$ , a focus is said to be *stable* or *unstable* respectively. The direction of motion along the spirals (clockwise or counterclockwise) is determined by the sign of the number  $b$ . In Fig. 5.3 are shown phase portraits of a stable focus ( $a < 0$ ) and an unstable focus ( $a > 0$ ) in the case  $b > 0$ , that is, the case in which the motion along the spirals is counterclockwise.

(d.2) If  $a = 0$ , then the singular point  $x_1 = x_2 = 0$  is called a *center*. Relationship (5.57) defines in this case a rotation of the vector  $\mathbf{x}_0$  through the angle  $bt$ . The phase curves are concentric circles with common center  $x_1 = x_2 = 0$  along which the motion is either clockwise or counterclockwise according to the sign of the number  $b$ . The phase portrait of a center (for the case  $b > 0$ ) is shown in Fig. 5.3.

# Chapter 6

## Quadratic and Bilinear Forms

### 6.1 Basic Definitions

**Definition 6.1** A *quadratic form* in  $n$  variables  $x_1, \dots, x_n$  is a homogeneous second-degree polynomial in these variables. Therefore, only terms of degree two enter into this polynomial; that is, the terms are monomials of the form  $\varphi_{ij}x_i x_j$  for all possible values of  $i, j = 1, \dots, n$ , and so the polynomial has the form

$$\psi(x_1, \dots, x_n) = \sum_{i,j=1}^n \varphi_{ij}x_i x_j. \quad (6.1)$$

We note that in expression (6.1), there are like terms, such as  $x_i x_j = x_j x_i$ . We shall decide later how to deal with them.

Of course, every quadratic form (6.1) can be viewed as a function of the vector  $\mathbf{x} = x_1 \mathbf{e}_1 + \dots + x_n \mathbf{e}_n$ , where  $\mathbf{e}_1, \dots, \mathbf{e}_n$  is some fixed basis of the vector space  $L$  of degree  $n$ . We shall write this as

$$\psi(\mathbf{x}) = \sum_{i,j=1}^n \varphi_{ij}x_i x_j. \quad (6.2)$$

The given definition of quadratic form obviously is compatible with the more general definition of form of arbitrary degree given in Sect. 3.8 (see p. 127). We recall that in that section, a *form* of degree  $k$  was defined as a function  $F(\mathbf{x})$  of the vector  $\mathbf{x} \in L$ , where  $F(\mathbf{x})$  is written as a homogeneous polynomial of degree  $k$  in coordinates  $x_1, \dots, x_n$  in some (and hence any) basis of this vector space. Thus for  $k = 2$ , we obtain the above definition of quadratic form.

By a change in coordinates, that is, by a choice of another basis of the space  $L$ , a quadratic form  $\psi(\mathbf{x})$  will be written as previously in the form (6.2) with some other coordinates  $\varphi_{ij}$ .

Quadratic forms have the property of being very similar to linear functions, and in the sequel, we shall unite the theory of quadratic forms with that of linear functions and transformations. The following notion will serve as a foundation for this.

**Definition 6.2** A function  $\varphi(\mathbf{x}, \mathbf{y})$  that assigns to two vectors  $\mathbf{x}, \mathbf{y} \in L$  a scalar value is called a *bilinear form* on  $L$  if it is linear in each of its arguments, that is, if for every fixed  $\tilde{\mathbf{y}} \in L$ , the function  $\varphi(\mathbf{x}, \tilde{\mathbf{y}})$  as a function of  $\mathbf{x}$  is linear on  $L$  and for each fixed  $\tilde{\mathbf{x}} \in L$ , the function  $\varphi(\tilde{\mathbf{x}}, \mathbf{y})$  as a function of  $\mathbf{y}$  is linear on  $L$ .

In other words, the following conditions must be satisfied for all vectors of the space  $L$  and scalars  $\alpha$ :

$$\begin{aligned}\varphi(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}) &= \varphi(\mathbf{x}_1, \mathbf{y}) + \varphi(\mathbf{x}_2, \mathbf{y}), \\ \varphi(\alpha \mathbf{x}, \mathbf{y}) &= \alpha \varphi(\mathbf{x}, \mathbf{y}), \\ \varphi(\mathbf{x}, \mathbf{y}_1 + \mathbf{y}_2) &= \varphi(\mathbf{x}, \mathbf{y}_1) + \varphi(\mathbf{x}, \mathbf{y}_2), \\ \varphi(\mathbf{x}, \alpha \mathbf{y}) &= \alpha \varphi(\mathbf{x}, \mathbf{y}).\end{aligned}\tag{6.3}$$

If the space  $L$  consists of rows, we have a special case of the notion of *multilinear* function, which was introduced in Sect. 2.7 (for  $m = 2$ ).

If  $\mathbf{e}_1, \dots, \mathbf{e}_n$  is some basis of  $L$ , then we can write

$$\mathbf{x} = x_1 \mathbf{e}_1 + \dots + x_n \mathbf{e}_n, \quad \mathbf{y} = y_1 \mathbf{e}_1 + \dots + y_n \mathbf{e}_n,$$

and using equations (6.3), we obtain a formula that expresses (in the chosen basis) the bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  in terms of the coordinates of the vectors  $\mathbf{x}$  and  $\mathbf{y}$ :

$$\varphi(\mathbf{x}, \mathbf{y}) = \sum_{i,j=1}^n \varphi_{ij} x_i y_j, \quad \text{where } \varphi_{ij} = \varphi(\mathbf{e}_i, \mathbf{e}_j).\tag{6.4}$$

In this case, the square matrix  $\Phi = (\varphi_{ij})$  is called the *matrix of the bilinear form*  $\varphi$  in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ . In the case that  $\mathbf{x}$  and  $\mathbf{y}$  are rows, this formulation represents a special way of writing an arbitrary multilinear function as introduced in Sect. 2.7 (Theorem 2.29).

The relationship (6.4) shows that the value of  $\varphi(\mathbf{x}, \mathbf{y})$  can be expressed in terms of the elements of the matrix  $\Phi$  and the coordinates of the vectors  $\mathbf{x}$  and  $\mathbf{y}$  in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , which means that a bilinear form, as a function of the arguments  $\mathbf{x}$  and  $\mathbf{y}$ , is completely defined by its matrix  $\Phi$ . This same formula shows that if we replace the argument  $\mathbf{y}$  in the bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  by  $\mathbf{x}$ , where  $\mathbf{x} = (x_1, \dots, x_n)$ , we obtain the quadratic form  $\psi(\mathbf{x}) = \varphi(\mathbf{x}, \mathbf{x})$ , and moreover, any quadratic form (6.1) can be obtained in this way; to do so, we need only choose a bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  with matrix  $\Phi = (\varphi_{ij})$  satisfying the condition  $\varphi(\mathbf{e}_i, \mathbf{e}_j) = \varphi_{ij}$ , where  $\varphi_{ij}$  are the coefficients from (6.1).

It is easily seen that the set of bilinear forms on a vector space  $L$  is itself a vector space if we define on it in a natural way the operations of addition of bilinear forms and multiplication by a scalar. Clearly, the null vector in such a space is the bilinear form that is identically equal to zero.

The connection between the notion of bilinear form and that of linear transformation is based on the following result, which uses the notion of dual space.



**Theorem 6.3** *There is an isomorphism between the space of bilinear forms  $\varphi$  on the vector space  $L$  and the space  $\mathfrak{L}(L, L^*)$  of linear transformations  $\mathcal{A} : L \rightarrow L^*$ .*

*Proof* Let  $\varphi(\mathbf{x}, \mathbf{y})$  be a bilinear form on  $L$ . Let us associate with it the linear transformation  $\mathcal{A} : L \rightarrow L^*$  as follows. By definition,  $\mathcal{A}$  should assign to a vector  $\mathbf{y} \in L$  a linear function  $\psi(\mathbf{x})$  on  $L$ . We shall make this assignment by setting  $\psi(\mathbf{x}) = \varphi(\mathbf{x}, \mathbf{y})$ . The verification that the transformation  $\mathcal{A}$  thus defined is linear is trivial.

It is equally trivial to verify that the correspondence  $\varphi \mapsto \mathcal{A}$  is a bijection. We shall simply point out the inverse transformation of the set  $\mathfrak{L}(L, L^*)$  into the set of bilinear forms. Let  $\mathcal{A}$  be a linear transformation from  $L$  to  $L^*$  that to each vector  $\mathbf{x} \in L$  assigns the linear function  $\mathcal{A}(\mathbf{x}) \in L^*$ . This function takes the value  $\mathcal{A}(\mathbf{x})(\mathbf{y})$  on the vector  $\mathbf{y}$ , which we shall denote by  $\varphi(\mathbf{x}, \mathbf{y})$ . Using the notation established in Sect. 3.7 (p. 125) and keeping in mind that in this situation,  $M = L^*$ , we may write  $\varphi(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathcal{A}(\mathbf{y}))$  for arbitrary vectors  $\mathbf{x}, \mathbf{y} \in L$ .

Finally, it is completely obvious that the constructed mapping  $\varphi \mapsto \mathcal{A}$  is an isomorphism of vector spaces, that is, it satisfies the conditions  $\varphi_1 + \varphi_2 \mapsto \mathcal{A}_1 + \mathcal{A}_2$  and  $\lambda\varphi \mapsto \lambda\mathcal{A}$ , where  $\varphi_i \mapsto \mathcal{A}_i$  and  $\lambda$  is an arbitrary scalar.  $\square$

It follows from this theorem that the study of bilinear forms is analogous to that of linear transformations  $L \rightarrow L$  (although somewhat simpler). In mathematics and physics, a special role is played by two particular types of bilinear form.

**Definition 6.4** A bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  is said to be *symmetric* if

$$\varphi(\mathbf{x}, \mathbf{y}) = \varphi(\mathbf{y}, \mathbf{x}), \quad (6.5)$$

and *antisymmetric* if

$$\varphi(\mathbf{x}, \mathbf{y}) = -\varphi(\mathbf{y}, \mathbf{x}), \quad (6.6)$$

for all vectors  $\mathbf{x}, \mathbf{y} \in L$ .

We encountered special cases of both these concepts in Chap. 2, when the vectors  $\mathbf{x}$  and  $\mathbf{y}$  were taken to be rows of numbers.

If following Theorem 6.3, we express the bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  in the form

$$\varphi(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathcal{A}(\mathbf{y})) \quad (6.7)$$

with some linear transformation  $\mathcal{A} : L \rightarrow L^*$ , then the symmetry condition (6.5) indicates that  $(\mathbf{x}, \mathcal{A}(\mathbf{y})) = (\mathbf{y}, \mathcal{A}(\mathbf{x}))$ . Since  $(\mathbf{y}, \mathcal{A}(\mathbf{x})) = (\mathbf{x}, \mathcal{A}^*(\mathbf{y}))$ , where  $\mathcal{A}^* : L^{**} \rightarrow L^*$  is the linear transformation dual to  $\mathcal{A}$  (see p. 125), then it can be rewritten in the form  $(\mathbf{x}, \mathcal{A}(\mathbf{y})) = (\mathbf{x}, \mathcal{A}^*(\mathbf{y}))$ . Since this relationship must be satisfied for all vectors  $\mathbf{x}, \mathbf{y} \in L$ , it can be rewritten in the form  $\mathcal{A} = \mathcal{A}^*$ . Note that in view of the equality  $L^{**} = L$ , both  $\mathcal{A}$  and  $\mathcal{A}^*$  are transformations from  $L$  to  $L^*$ . Similarly, the antisymmetry condition (6.6) of the bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  can be written in the form  $\mathcal{A} = -\mathcal{A}^*$ .

Let us note that it suffices to verify the symmetry condition (6.5) and antisymmetry condition (6.6) for vectors  $\mathbf{x}$  and  $\mathbf{y}$  belonging to some particular basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the space  $L$ . Indeed, if this condition is satisfied for vectors in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , that is, for example, in the case of symmetry, the equations  $\varphi(\mathbf{e}_i, \mathbf{e}_j) = \varphi(\mathbf{e}_j, \mathbf{e}_i)$  are satisfied for all  $i, j = 1, \dots, n$ , then from formula (6.4), it follows that the condition (6.5) is met for all vectors  $\mathbf{x}, \mathbf{y} \in L$ . Recalling the definition of a matrix of a bilinear form, we see that the form  $\varphi$  is symmetric if and only if its matrix  $\Phi$  is symmetric in some basis of the space  $L$  (that is,  $\Phi = \Phi^*$ ). Similarly, the antisymmetry of the bilinear form  $\varphi$  is equivalent to the antisymmetry of  $\Phi$  in some basis ( $\Phi = -\Phi^*$ ).

The matrix  $\Phi$  of a bilinear form depends on the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ . We shall now investigate this dependence. Here, we shall use the formula (3.38) for changing coordinates that we derived in Sect. 3.4, and moreover, our reasoning will be similar to what we used then in deriving this formula.

First of all, let us write down the relationship (6.4) in a more compact matrix form. To this end, we observe that for

$$\text{rows } \mathbf{x} = (x_1, \dots, x_n) \quad \text{and} \quad \text{columns } [\mathbf{y}] = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

the sum in formula (6.4) can be rewritten in the following form:

$$\sum_{i,j=1}^n \varphi_{ij} x_i y_j = \sum_{i=1}^n x_i \left( \sum_{j=1}^n \varphi_{ij} y_j \right) = \sum_{i=1}^n x_i z_i, \quad \text{where } z_i = \sum_{j=1}^n \varphi_{ij} y_j.$$

By the rule of matrix multiplication, we obtain the expression

$$\sum_{i,j=1}^n \varphi_{ij} x_i y_j = \mathbf{x}[\mathbf{z}], \quad \text{where } [\mathbf{z}] = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} = \Phi[\mathbf{y}].$$

This means that we now have

$$\sum_{i,j=1}^n \varphi_{ij} x_i y_j = \mathbf{x} \Phi[\mathbf{y}].$$

Let us note that by similar arguments, or by simply taking the transpose of both sides of the previous equality (on the left-hand side of which stands a scalar, that is, a matrix of type  $(1, 1)$ , which is invariant under the transpose operation), we obtain a similar relationship

$$\sum_{i,j=1}^n \varphi_{ij} x_i y_j = \mathbf{y} \Phi^*[\mathbf{x}].$$

Thus if in some basis  $e_1, \dots, e_n$ , the matrix of the bilinear form  $\varphi$  is equal to  $\Phi$ , while the vectors  $x$  and  $y$  have coordinates  $x_i$  and  $y_i$ , then we have the following formula:

$$\varphi(x, y) = x\Phi[y]. \quad (6.8)$$

Similarly, for another basis  $e'_1, \dots, e'_n$ , we obtain the equality

$$\varphi(x, y) = x'\Phi'[y'], \quad (6.9)$$

where  $\Phi'$  is the matrix of the bilinear form  $\varphi$ , while  $x'_i$  and  $y'_i$  are the coordinates of the vectors  $x$  and  $y$  in the basis  $e'_1, \dots, e'_n$ .

Let  $C$  be the transition matrix from the basis  $e'_1, \dots, e'_n$  to the basis  $e_1, \dots, e_n$ . Then by the substitution formula (3.36), we obtain the relationships  $x = x'C^*$  and  $[y] = C[y']$ . Substituting these expressions into (6.8), taking into account formula (6.9), we obtain the identity

$$x'C^*\Phi C[y'] = x'\Phi'[y'],$$

which is satisfied for all  $x'$  and  $[y']$ . From this, it follows that the matrices  $\Phi$  and  $\Phi'$  of the bilinear form  $\varphi$  in these bases are related by the equality

$$\Phi' = C^*\Phi C. \quad (6.10)$$

This is the substitution formula for the matrix of a bilinear form for a change of basis.

Since the rank of a matrix is invariant under multiplication on the left or right by a nonsingular square matrix of appropriate order (Theorem 2.63), it follows that the rank of the matrix  $\Phi$  is the same as that of the matrix  $\Phi'$  for any transition matrix  $C$ . Thus the rank  $r$  of the matrix of a bilinear form does not depend on the basis in which the matrix is written, and consequently, we may call it simply the *rank of the bilinear form*  $\varphi$ . In particular, if  $r = n$ , that is, if the rank coincides with the dimension of the vector space  $L$ , then the bilinear form  $\varphi$  is said to be *nonsingular*.

The rank of a bilinear form can be defined in another way. By Theorem 6.3, to every bilinear form  $\varphi$  there corresponds a unique linear transformation  $\mathcal{A} : L \rightarrow L^*$ , and the connection between the two is laid out in (6.7). It is easily verified that if we choose in the spaces  $L$  and  $L^*$  two dual bases, then the matrices of the bilinear form  $\varphi$  and the linear transformation  $\mathcal{A}$  will coincide. This shows that the rank of the bilinear form is the same as the rank of the linear transformation  $\mathcal{A}$ . From this we derive that in particular, the form  $\varphi$  is nonsingular if and only if the linear transformation  $\mathcal{A} : L \rightarrow L^*$  is an isomorphism.

A given quadratic form  $\psi$  can be obtained from different bilinear forms  $\varphi$ ; this is related to the presence of similar terms in the expression (6.1) for a quadratic form, about which we spoke above. In order to obtain uniqueness and agreement with the properties of linearity, we shall proceed not as in secondary school, where, for example, one writes the sum of terms  $a_{12}x_1x_2 + a_{21}x_2x_1 = (a_{12} + a_{21})x_1x_2$ , but instead using a notation in which we do not collect like terms.

**Remark 6.5** (On elements of fields) Additional refinements in this section are directed at the reader who is interested in the case of vector spaces over an arbitrary field  $\mathbb{K}$ . Here we shall introduce a certain limitation that will allow us to provide a single account for the cases  $\mathbb{K} = \mathbb{R}$ ,  $\mathbb{K} = \mathbb{C}$ , and all types of fields that we will be concerned with. Namely, in what follows we shall assume that  $\mathbb{K}$  is a *field of characteristic different from 2*.<sup>1</sup> (We mentioned a similar limitation in the general concept of field on p. 83.) Using the simplest properties that can be derived from the definition of a field, it is easy to prove that in a field of characteristic different from 2, there exists for an arbitrary element  $a$  a unique element  $b$  such that  $2b = a$  (where  $2b$  denotes the sum  $b + b$ ). We then set  $b = a/2$ , and so whenever  $a = 0$ , it follows that  $b = 0$ .

**Theorem 6.6** *Every quadratic form  $\psi(\mathbf{x})$  on the space  $L$  over a field  $\mathbb{K}$  of characteristic different from 2 can be represented in the form*

$$\psi(\mathbf{x}) = \varphi(\mathbf{x}, \mathbf{x}), \quad (6.11)$$

where  $\varphi$  is a symmetric bilinear form, and moreover, for the given quadratic form  $\psi$ , the bilinear form  $\varphi$  is unique.

*Proof* By what we have said above, an arbitrary quadratic form  $\psi(\mathbf{x})$  can be represented in the form

$$\psi(\mathbf{x}) = \varphi_1(\mathbf{x}, \mathbf{x}), \quad (6.12)$$

where  $\varphi_1(\mathbf{x}, \mathbf{y})$  is some bilinear form, not necessarily symmetric. Let us set

$$\varphi(\mathbf{x}, \mathbf{y}) = \frac{\varphi_1(\mathbf{x}, \mathbf{y}) + \varphi_1(\mathbf{y}, \mathbf{x})}{2}.$$

It is clear that  $\varphi(\mathbf{x}, \mathbf{y})$  is a bilinear form, and moreover, it is already symmetric. From formula (6.12) follows the relationship (6.11), as asserted.

We shall now prove that if relationship (6.11) holds for some symmetric bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$ , then  $\varphi(\mathbf{x}, \mathbf{y})$  is uniquely determined by the quadratic form  $\psi(\mathbf{x})$ . To see this, let us calculate  $\psi(\mathbf{x} + \mathbf{y})$ . By assumption and the properties of the bilinear form  $\varphi$ , we have

$$\psi(\mathbf{x} + \mathbf{y}) = \varphi(\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) = \varphi(\mathbf{x}, \mathbf{x}) + \varphi(\mathbf{y}, \mathbf{y}) + \varphi(\mathbf{x}, \mathbf{y}) + \varphi(\mathbf{y}, \mathbf{x}). \quad (6.13)$$

In view of the symmetry of the form  $\varphi$ , we have

$$\psi(\mathbf{x} + \mathbf{y}) = \psi(\mathbf{x}) + \psi(\mathbf{y}) + 2\varphi(\mathbf{x}, \mathbf{y}),$$

---

<sup>1</sup>Fields of characteristic different from 2 are what are most frequently encountered. However, fields of characteristic 2, which we are excluding from consideration here, have important applications, for example in discrete mathematics and cryptography.

which implies that

$$\varphi(\mathbf{x}, \mathbf{y}) = \frac{1}{2}(\psi(\mathbf{x} + \mathbf{y}) - \psi(\mathbf{x}) - \psi(\mathbf{y})). \quad (6.14)$$

This last relationship uniquely determines a bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  associated with the given quadratic form  $\psi(\mathbf{x})$ .  $\square$

With the same assumptions, we have the following result for antisymmetric forms.

**Theorem 6.7** *For every antisymmetric bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  on the space  $\mathbb{L}$  over a field  $\mathbb{K}$  of characteristic different from 2, we have*

$$\varphi(\mathbf{x}, \mathbf{x}) = 0. \quad (6.15)$$

*Conversely, if equality (6.15) is satisfied for every vector  $\mathbf{x} \in \mathbb{L}$ , then the bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  is antisymmetric.*

*Proof* If the form  $\varphi(\mathbf{x}, \mathbf{y})$  is antisymmetric, then transposing the arguments in the expression  $\varphi(\mathbf{x}, \mathbf{x})$  leads to the relationship  $\varphi(\mathbf{x}, \mathbf{x}) = -\varphi(\mathbf{x}, \mathbf{x})$ , and then  $2\varphi(\mathbf{x}, \mathbf{x}) = 0$ , from which follows equality (6.15), since by the condition of the theorem, the field  $\mathbb{K}$  has characteristic different from 2. Conversely, if  $\varphi(\mathbf{x}, \mathbf{x}) = 0$  for every vector  $\mathbf{x} \in \mathbb{L}$ , then this holds in particular for the vector  $\mathbf{x} + \mathbf{y}$ , that is, we obtain

$$\varphi(\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) = \varphi(\mathbf{x}, \mathbf{x}) + \varphi(\mathbf{x}, \mathbf{y}) + \varphi(\mathbf{y}, \mathbf{x}) + \varphi(\mathbf{y}, \mathbf{y}) = 0.$$

Since we have  $\varphi(\mathbf{x}, \mathbf{x}) = \varphi(\mathbf{y}, \mathbf{y}) = 0$  by the hypothesis of the theorem, it follows that  $\varphi(\mathbf{x}, \mathbf{y}) + \varphi(\mathbf{y}, \mathbf{x}) = 0$ , which yields that the bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  is antisymmetric.  $\square$

Let us note that the way of writing the quadratic form  $\psi(\mathbf{x})$  in the form (6.11) established by Theorem 6.6, where  $\varphi(\mathbf{x}, \mathbf{y})$  is a symmetric bilinear form, shows us how to write similar terms in the representation (6.1). Indeed, if we have

$$\mathbf{x} = x_1 \mathbf{e}_1 + \cdots + x_n \mathbf{e}_n, \quad \mathbf{y} = y_1 \mathbf{e}_1 + \cdots + y_n \mathbf{e}_n,$$

and  $\varphi(\mathbf{x}, \mathbf{y})$  is a bilinear form, then

$$\varphi(\mathbf{x}, \mathbf{y}) = \sum_{i,j=1}^n \varphi_{ij} x_i y_j,$$

where  $\varphi_{ij} = \varphi(\mathbf{e}_i, \mathbf{e}_j)$ . The symmetry of the form  $\varphi(\mathbf{x}, \mathbf{y})$  implies that  $\varphi_{ij} = \varphi_{ji}$  for all  $i, j = 1, \dots, n$ . Then the representation

$$\psi(x_1, \dots, x_n) = \sum_{i,j=1}^n \varphi_{ij} x_i x_j$$

contains like terms  $\varphi_{ij}x_i x_j$  and  $\varphi_{ji}x_j x_i$  for  $i \neq j$ . Then if  $i \neq j$ , the term with  $x_i x_j$  occurs in the sum twice: as  $\varphi_{ij}x_i x_j$  and as  $\varphi_{ji}x_j x_i$ . Since  $\varphi_{ij} = \varphi_{ji}$ , then collecting like terms leads to this sum being written in the form  $2\varphi_{ij}x_i x_j$ .

For example, the coefficients of the quadratic form  $x_1^2 + x_1 x_2 + x_2^2$  are given by  $\varphi_{11} = 1$ ,  $\varphi_{22} = 1$ , and  $\varphi_{12} = \varphi_{21} = \frac{1}{2}$ . Such a way of writing things may seem strange at first glance, but as we shall soon see, it offers many advantages.

## 6.2 Reduction to Canonical Form

The main goal of this section is to transform quadratic forms into the simplest possible form, called *canonical*. As in the case of the matrix of a linear transformation, canonical form is obtained by the selection of a special basis of the given vector space. Namely, the required basis must possess the property that the matrix of the symmetric bilinear form corresponding to the given quadratic form assumes diagonal form in that basis. This property is directly connected to the important notion of *orthogonality*, which will be used repeatedly in this and subsequent chapters. We note that the notion of orthogonality can be formulated in a way that is well defined for bilinear forms that are not necessarily symmetric, but it can be most simply defined for symmetric and antisymmetric bilinear forms. In this section, we shall consider only symmetric bilinear forms.

Thus let there be given on the finite-dimensional vector space  $L$  a symmetric bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$ .

**Definition 6.8** Vectors  $\mathbf{x}$  and  $\mathbf{y}$  are said to be *orthogonal* if  $\varphi(\mathbf{x}, \mathbf{y}) = 0$ .

We observe that in light of the symmetry condition  $\varphi(\mathbf{y}, \mathbf{x}) = \varphi(\mathbf{x}, \mathbf{y})$ , the equality  $\varphi(\mathbf{x}, \mathbf{y}) = 0$  is equivalent to  $\varphi(\mathbf{y}, \mathbf{x}) = 0$ . This is true as well for antisymmetric bilinear forms. However, if we do not impose a symmetry or antisymmetry condition on the bilinear form, then the vector  $\mathbf{x}$  can be orthogonal to the vector  $\mathbf{y}$  without  $\mathbf{y}$  being orthogonal to  $\mathbf{x}$ . This leads to the concepts of left and right orthogonality and some very beautiful geometry, but it would take us beyond the scope of this book. A vector  $\mathbf{x} \in L$  is said to be *orthogonal* to a subspace  $L' \subset L$  relative to  $\varphi$  if it is orthogonal to every vector  $\mathbf{y} \in L'$ , that is, if  $\varphi(\mathbf{x}, \mathbf{y}) = 0$  for all  $\mathbf{y} \in L'$ .

It follows at once from the definition of bilinearity that the collection of all vectors  $\mathbf{x}$  orthogonal to a subspace  $L'$  with respect to a given bilinear form  $\varphi$  is itself a subspace of  $L$ . It is called the *orthogonal complement* of the subspace  $L'$  with respect to the form  $\varphi$  and is denoted by  $(L')_{\varphi}^{\perp}$ .

In particular, for  $L' = L$ , the subspace  $(L)_{\varphi}^{\perp}$  represents the totality of vectors  $\mathbf{x} \in L$  for which the equation  $\varphi(\mathbf{x}, \mathbf{y}) = 0$  is satisfied for all  $\mathbf{y} \in L$ . This subspace is called the *radical* of the bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$ . From the definition of a bilinear form, it follows at once that the radical consists of all vectors  $\mathbf{x} \in L$  such that

$$\varphi(\mathbf{x}, \mathbf{e}_i) = 0 \quad \text{for all } i = 1, \dots, n, \quad (6.16)$$

where  $e_1, \dots, e_n$  is some basis of the space  $L$ . The equalities (6.16) are linear homogeneous equations that define the radical as a subspace of  $L$ . If we write down the vector  $x$  in the chosen basis, that is, in the form  $x = x_1 e_1 + \dots + x_n e_n$ , then in view of formula (6.4), we obtain from the equalities (6.16) a system of linear homogeneous equations in the unknowns  $x_1, \dots, x_n$ . The matrix of this system coincides with the matrix  $\Phi$  of the bilinear form  $\varphi$  in the basis  $e_1, \dots, e_n$ . Thus the space  $(L)_\varphi^\perp$  satisfies the conditions of Example 3.65 from Sect. 3.5 (p. 114). Consequently,  $\dim(L)_\varphi^\perp = n - r$ , where  $r$  is the rank of the matrix of the linear system, that is, the rank of the bilinear form  $\varphi$ . We therefore obtain the equality

$$r = \dim L - \dim(L)_\varphi^\perp. \quad (6.17)$$

**Theorem 6.9** *Let  $L' \subset L$  be a subspace such that the restriction of the bilinear form  $\varphi(x, y)$  to  $L'$  is a nonsingular bilinear form. We then have the decomposition*

$$L = L' \oplus (L')_\varphi^\perp. \quad (6.18)$$

*Proof* First of all, we note that by the conditions of the theorem, the intersection  $L' \cap (L')_\varphi^\perp$  is equal to the zero space  $(0)$ . Indeed, it consists of all vectors  $x \in L'$  such that  $\varphi(x, y) = 0$  for all  $y \in L'$ , and hence only for the null vector, since by the condition, the restriction of  $\varphi$  to the subspace  $L'$  is a nonsingular bilinear form. Thus it suffices to prove that  $L' + (L')_\varphi^\perp = L$ . We shall present two proofs of this fact in order to demonstrate two different lines of reasoning used in the theory of vector spaces.

**First proof.** We shall use the linear transformation  $\mathcal{A} : L \rightarrow L^*$  constructed in Theorem 6.3 corresponding to the bilinear form  $\varphi$ . Assigning to each linear function on  $L$  its restriction to the subspace  $L' \subset L$ , we obtain the linear transformation  $\mathcal{B} : L^* \rightarrow (L')^*$ . If we apply in sequence the linear transformations  $\mathcal{A}$  and  $\mathcal{B}$ , we obtain the linear transformation  $\mathcal{C} = \mathcal{B}\mathcal{A} : L \rightarrow (L')^*$ .

The kernel  $L_1$  of the transformation  $\mathcal{C}$  consists of the vectors  $y \in L$  such that  $\varphi(x, y) = 0$  for all  $x \in L'$ , since by definition,  $\varphi(x, y) = (x, \mathcal{A}(y))$ . This implies that  $L_1 = (L')_\varphi^\perp$ . Let us show that the image  $L_2$  of the transformation  $\mathcal{C}$  is equal to the entire subspace  $(L')^*$ . We shall prove an even stronger result: an arbitrary vector  $u \in (L')^*$  can be represented in the form  $u = \mathcal{C}(v)$ , where  $v \in L'$ . For this, we must consider the restriction of the transformation  $\mathcal{C}$  to the subspace  $L'$ . By definition, it coincides with the transformation  $\mathcal{A}' : L' \rightarrow (L')^*$  constructed in Theorem 6.3, which corresponds to the restriction of the bilinear form  $\varphi$  to  $L'$ . By assumption, the restriction of the form  $\varphi$  to  $L'$  is nonsingular, which implies that the transformation  $\mathcal{A}'$  is an isomorphism. From this, it follows in particular that its image is the entire subspace  $(L')^*$ .

Now we shall make use of Theorem 3.72 and apply relationship (3.47) to the transformation  $\mathcal{C}$ . We obtain  $\dim L_1 + \dim L_2 = \dim L$ . Since  $L_2 = (L')^*$ , it follows by Theorem 3.78 that  $\dim L_2 = \dim L'$ . Recalling also that  $L_1 = (L')_\varphi^\perp$ , we have finally the equality

$$\dim(L')_\varphi^\perp + \dim L' = \dim L. \quad (6.19)$$

Since  $L' \cap (L')_{\varphi}^{\perp} = \{\mathbf{0}\}$ , we conclude by Corollary 3.15 (p. 85) that  $L' + (L')_{\varphi}^{\perp} = L' \oplus (L')_{\varphi}^{\perp}$ . From Theorems 3.24, 3.38 and the relationship (6.19), it follows that  $L' \oplus (L')_{\varphi}^{\perp} = L$ .

**Second proof.** We need to represent an arbitrary vector  $\mathbf{x} \in L$  in the form  $\mathbf{x} = \mathbf{u} + \mathbf{v}$ , where  $\mathbf{u} \in L'$  and  $\mathbf{v} \in (L')_{\varphi}^{\perp}$ . This is clearly equivalent to the condition  $\mathbf{x} - \mathbf{u} \in (L')_{\varphi}^{\perp}$ , and therefore to the condition  $\varphi(\mathbf{x} - \mathbf{u}, \mathbf{y}) = 0$  for all  $\mathbf{y} \in L'$ . Recalling the properties of a bilinear form, we see that it suffices that the last equation be satisfied for vectors  $\mathbf{y} = \mathbf{e}_i$ ,  $i = 1, \dots, r$ , where  $\mathbf{e}_1, \dots, \mathbf{e}_r$  is some basis of the space  $L'$ . In view of the bilinearity of the form  $\varphi$ , our relationships can be written in the form

$$\varphi(\mathbf{u}, \mathbf{e}_i) = \varphi(\mathbf{x}, \mathbf{e}_i) \quad \text{for all } i = 1, \dots, r. \quad (6.20)$$

We now represent the vector  $\mathbf{u}$  as  $\mathbf{u} = x_1 \mathbf{e}_1 + \dots + x_r \mathbf{e}_r$ . Relationship (6.20) gives a system of  $r$  linear equations

$$\varphi(\mathbf{e}_1, \mathbf{e}_i)x_1 + \dots + \varphi(\mathbf{e}_r, \mathbf{e}_i)x_r = \varphi(\mathbf{x}, \mathbf{e}_i), \quad i = 1, \dots, r, \quad (6.21)$$

with unknowns  $x_1, \dots, x_r$ . The matrix of the system (6.21) has the form

$$\Phi = \begin{pmatrix} \varphi(\mathbf{e}_1, \mathbf{e}_1) & \dots & \varphi(\mathbf{e}_1, \mathbf{e}_r) \\ \vdots & \ddots & \vdots \\ \varphi(\mathbf{e}_r, \mathbf{e}_1) & \dots & \varphi(\mathbf{e}_r, \mathbf{e}_r) \end{pmatrix}.$$

But it is easy to see that  $\Phi$  is the matrix of the restriction of the bilinear form  $\varphi$  to the subspace  $L'$  written in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_r$ . Since by assumption, such a form is nonsingular, its matrix is also nonsingular, and this implies that the system of equations (6.20) has a solution. In other words, we can find a vector  $\mathbf{u} \in L'$  satisfying all the relationships (6.20), which proves our assertion.  $\square$

We shall now apply these ideas related to bilinear forms to the theory of quadratic forms. Our goal is to find a basis in which the matrix of a given quadratic form  $\psi(\mathbf{x})$  has the simplest form possible.

**Theorem 6.10** *For every quadratic form  $\psi(\mathbf{x})$ , there exists a basis in which the form can be written as*

$$\psi(\mathbf{x}) = \lambda_1 x_1^2 + \dots + \lambda_n x_n^2, \quad (6.22)$$

where  $x_1, \dots, x_n$  are the coordinates of the vector  $\mathbf{x}$  in this basis.

*Proof* Let  $\varphi(\mathbf{x}, \mathbf{y})$  be a symmetric bilinear form associated with the quadratic form  $\psi(\mathbf{x})$  by the formula (6.11). If  $\psi(\mathbf{x})$  is identically equal to zero, then the theorem clearly is true (for  $\lambda_1 = \dots = \lambda_n = 0$ ). If the quadratic form  $\psi(\mathbf{x})$  is not identically equal to zero, then there exists a vector  $\mathbf{e}_1$  such that  $\psi(\mathbf{e}_1) \neq 0$ , that is,  $\varphi(\mathbf{e}_1, \mathbf{e}_1) \neq 0$ . This implies that the restriction of the bilinear form  $\varphi$  to the subspace  $L' = \langle \mathbf{e}_1 \rangle$  is



nonsingular, and therefore, by Theorem 6.9, for the subspace  $L' = \langle \mathbf{e}_1 \rangle$  we have the decomposition (6.18), that is,  $L = \langle \mathbf{e}_1 \rangle \oplus \langle \mathbf{e}_1 \rangle_\phi^\perp$ . Since  $\dim \langle \mathbf{e}_1 \rangle = 1$ , then by Theorem 3.38, we obtain that  $\dim \langle \mathbf{e}_1 \rangle_\phi^\perp = n - 1$ .

Proceeding by induction, we may assume the theorem to have been proved for the space  $\langle \mathbf{e}_1 \rangle_\phi^\perp$ . Thus in this space there exists a basis  $\mathbf{e}_2, \dots, \mathbf{e}_n$  such that  $\varphi(\mathbf{e}_i, \mathbf{e}_j) = 0$  for all  $i \neq j$ ,  $i, j \geq 2$ . Then in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the space  $L$ , the quadratic form  $\psi(\mathbf{x})$  can be written as (6.22) for some  $\lambda_1, \dots, \lambda_n$ .  $\square$

We observe that one and the same quadratic form  $\psi$  can be of the form (6.22) in various bases, and in this case, the numbers  $\lambda_1, \dots, \lambda_n$  might differ in various bases. For example, if in a one-dimensional space whose basis consists of one nonzero vector  $\mathbf{e}$ , we define the quadratic form  $\psi$  by the relation  $\psi(x\mathbf{e}) = x^2$ , then in the basis consisting of the vector  $\mathbf{e}' = \lambda\mathbf{e}$ ,  $\lambda \neq 0$ , it can be written as  $\psi(x\mathbf{e}') = (\lambda x)^2$ .

If in a certain basis a quadratic form can be written as in (6.22), then we say that in that basis, it is in *canonical form*. Theorem 6.10 is called the *theorem on reducing a quadratic form to canonical form*. From what we have said above, it follows that reducing a quadratic form to canonical form is not unique.

If in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the space  $L$ , the quadratic form  $\psi(\mathbf{x})$  has the form established in Theorem 6.10, then its matrix in this basis is equal to

$$\Psi = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}. \quad (6.23)$$

It is clear that the rank of the matrix  $\Psi$  is equal to the number of nonzero values among  $\lambda_1, \dots, \lambda_n$ . As we saw in the previous section, the rank of the matrix  $\Psi$  (that is, the rank of the quadratic form  $\psi(\mathbf{x})$ ) does not depend on the choice of basis in which the matrix  $\Psi$  is written. Therefore, this number is the same for every basis for which Theorem 6.10 holds.

It is useful to write down the results we have obtained in matrix form. We may reformulate Theorem 6.10 using formula (6.10) obtained in the previous section for replacing the matrix of a bilinear form by a change in basis.

**Theorem 6.11** *For an arbitrary symmetric matrix  $\Phi$ , there exists a nonsingular matrix  $C$  such that the matrix  $C^* \Phi C$  is diagonal. If we select a different matrix  $C$ , we may obtain different diagonal matrices  $C^* \Phi C$ , but the number of nonzero elements on the main diagonal will always be the same.*

A completely analogous argument can be applied to the case of antisymmetric bilinear forms. The following theorem is an analogue of Theorem 6.10.

**Theorem 6.12** *For every antisymmetric bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$ , there exists a basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  whose first  $2r$  vectors can be combined into pairs  $(\mathbf{e}_{2i-1}, \mathbf{e}_{2i})$ ,  $i =$*

$1, \dots, r$ , such that

$$\begin{aligned}\varphi(\mathbf{e}_{2i-1}, \mathbf{e}_{2i}) &= 1, & \varphi(\mathbf{e}_{2i}, \mathbf{e}_{2i-1}) &= -1 \quad \text{for all } i = 1, \dots, r, \\ \varphi(\mathbf{e}_i, \mathbf{e}_j) &= 0 \quad \text{if } |i - j| > 1 \text{ or } i > 2r \text{ or } j > 2r.\end{aligned}$$

Thus in the given basis, the matrix of the bilinear form  $\varphi$  takes the form

$$\Phi = \begin{pmatrix} 0 & 1 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ -1 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & 0 & 1 & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & -1 & 0 & \dots & \dots & \dots & \dots & \dots & \dots \\ & & & & \ddots & & & & & \\ \dots & \dots & \dots & \dots & \dots & 0 & 1 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & -1 & 0 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 & \dots & \dots \\ & & & & & & & & \ddots & \\ 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 \end{pmatrix}. \quad (6.24)$$

*Proof* This theorem is an exact parallel to Theorem 6.10. If  $\varphi(\mathbf{x}, \mathbf{y}) = 0$  for all  $\mathbf{x}$  and  $\mathbf{y}$ , then the assertion of the theorem is obvious (for  $r = 0$ ). However, if this is not the case, then there exist two vectors  $\mathbf{e}'_1$  and  $\mathbf{e}_2$  for which  $\varphi(\mathbf{e}'_1, \mathbf{e}_2) = \alpha \neq 0$ . Setting  $\mathbf{e}_1 = \alpha^{-1}\mathbf{e}'_1$ , we obtain that  $\varphi(\mathbf{e}_1, \mathbf{e}_2) = 1$ . The matrix of the form  $\varphi$  restricted to the subspace  $L' = \langle \mathbf{e}_1, \mathbf{e}_2 \rangle$  in the basis  $\mathbf{e}_1, \mathbf{e}_2$  has the form

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad (6.25)$$

and consequently, it is nonsingular. Then on the basis of Theorem 6.9, we obtain the decomposition  $L = L' \oplus (L')^\perp_\varphi$ , where  $\dim(L')^\perp_\varphi = n - 2$ , with  $n = \dim L$ . Proceeding by induction, we may assume that the theorem has been proved for forms  $\varphi$  defined on the space  $(L')^\perp_\varphi$ . If  $\mathbf{f}_1, \dots, \mathbf{f}_{n-2}$  is such a basis of the space  $(L')^\perp_\varphi$ , the existence of which is asserted by Theorem 6.12, then it is obvious that  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{f}_1, \dots, \mathbf{f}_{n-2}$  is the required basis of the original space  $L$ .  $\square$

The number  $n - 2r$  is equal to the dimension of the radical of the bilinear form  $\varphi$ , and therefore, it is the same for all bases in which the matrix of the bilinear form  $\varphi$  is brought into the form (6.24). The rank of the matrix (6.25) is equal to 2, while the matrix (6.24) contains  $r$  such blocks on the main diagonal. Therefore, the rank of the matrix (6.24) is equal to  $2r$ . Thus from Theorem 6.12, we obtain the following corollary.

**Corollary 6.13** *The rank of an antisymmetric bilinear form is an even number.*

Let us now translate everything that we have proved for antisymmetric bilinear forms into the language of matrices. Here our assertions will be the same as for

symmetric matrices, and they are proved in exactly the same manner. We obtain that for an arbitrary antisymmetric matrix  $\Phi$ , there exists a nonsingular matrix  $C$  such that the matrix

$$\Phi' = C^* \Phi C \quad (6.26)$$

has the form (6.24).

Matrices  $\Phi$  and  $\Phi'$  that are related by (6.26) for some nonsingular matrix  $C$  are said to be *equivalent*. The same term is applied to the quadratic forms associated with these matrices (for a particular choice of basis).

It is easy to verify that the concept thus introduced is an equivalence relation on the set of square matrices of a given order or indeed on the set of quadratic forms. The reflexive property is obvious. It is necessary only to substitute the matrix  $C = E$  into formula (6.26). Multiplying both sides of equality (6.26) on the right by the matrix  $B = C^{-1}$  and on the left by the matrix  $B^*$ , taking into account the relationship  $(C^{-1})^* = (C^*)^{-1}$ , we obtain the equality  $\Phi = B^* \Phi' B$ , which establishes the symmetric property.

Finally, let us verify the property of transitivity. Suppose we are given the relationships (6.26) and  $\Phi'' = D^* \Phi' D$  for some nonsingular matrices  $C$  and  $D$ . Then if we substitute the first of these into the second, we obtain the equality  $\Phi'' = D^* C^* \Phi C D$ . Setting  $B = C D$  and taking into account  $B^* = D^* C^*$ , we obtain the equality  $\Phi'' = B^* \Phi B$ , which establishes the equivalence of the matrices  $\Phi$  and  $\Phi''$ .

It is now possible to reformulate Theorems 6.10 and 6.12 in the following form.

**Theorem 6.14** *Every symmetric matrix is equivalent to a diagonal matrix.*

**Theorem 6.15** *Every antisymmetric matrix  $\Phi$  is equivalent to a matrix of the form (6.24), where the number  $r$  is equal to one-half the rank of the matrix  $\Phi$ .*

From Theorems 6.14 and 6.15, it follows that all equivalent symmetric matrices and all equivalent antisymmetric matrices have the same rank, and for antisymmetric matrices, equivalence is the same as the equality of their ranks, that is, two antisymmetric matrices of a given order are equivalent if and only if they have the same rank.

Let us conclude with the observation that all the concepts investigated in this section can be expressed in the language of bilinear forms given by Theorem 6.3. By this theorem, every bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  on a vector space  $L$  can be written uniquely in the form  $\varphi(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathcal{A}(\mathbf{y}))$ , where  $\mathcal{A} : L \rightarrow L^*$  is some linear transformation. As proved in Sect. 6.1, the symmetry of the form  $\varphi$  is equivalent to  $\mathcal{A}^* = \mathcal{A}$ , while antisymmetry is equivalent to  $\mathcal{A}^* = -\mathcal{A}$ . In the first case, the transformation  $\mathcal{A}$  is said to be *symmetric*, and in the second case, *antisymmetric*. Thus Theorems 6.10 and 6.12 are equivalent to the following assertions. For an arbitrary symmetric transformation  $\mathcal{A}$ , there exists a basis of the vector space  $L$  in which the matrix of this transformation has the diagonal form (6.23). Similarly, for an arbitrary antisymmetric transformation  $\mathcal{A}$ , there exists a basis of the space  $L$  in which the matrix of this

transformation has the form (6.24). More precisely, in both these statements, we are talking about the choice of basis in  $L$  and its dual basis in  $L^*$ , since the transformation  $\mathcal{A}$  maps  $L$  to  $L^*$ .

### 6.3 Complex, Real, and Hermitian Forms

We begin this section by examining a quadratic form  $\psi$  in a complex vector space  $L$ . By Theorem 6.10, it can be written, in terms of some basis  $e_1, \dots, e_n$ , in the form

$$\psi(x) = \lambda_1 x_1^2 + \dots + \lambda_n x_n^2,$$

where  $x_1, \dots, x_n$  are the coordinates of the vector  $x$  in this basis. This implies that for the associated symmetric bilinear form  $\varphi(x, y)$ , it has the value  $\varphi(e_i, e_j) = 0$  for  $i \neq j$  and  $\varphi(e_i, e_i) = \lambda_i$ . Here, the number of values  $\lambda_i$  different from zero is equal to the rank  $r$  of the bilinear form  $\varphi$ . By changing the numeration of the basis vectors if necessary, we may assume that  $\lambda_i \neq 0$  for  $i \leq r$  and  $\lambda_i = 0$  for  $i > r$ . We may then introduce a new basis  $e'_1, \dots, e'_n$  by setting

$$e'_i = \sqrt{\lambda_i} e_i \quad \text{for } i \leq r, \quad e'_i = e_i \quad \text{for } i > r,$$

since  $\sqrt{\lambda_i}$  is again a complex number. In the new basis,  $\varphi(e'_i, e'_j) = 0$  for all  $i \neq j$  and  $\varphi(e'_i, e'_i) = 1$  for  $i \leq r$ ,  $\varphi(e'_i, e'_i) = 0$  for  $i > r$ . This implies that the quadratic form  $\psi(x)$  can be written in this basis in the form

$$\psi(x) = x_1^2 + \dots + x_r^2, \quad (6.27)$$

where  $x_1, \dots, x_r$  are the first  $r$  coordinates of the vector  $x$ . We see, then, that in a complex space  $L$ , every quadratic form can be brought into the canonical form (6.27), and all quadratic forms (and therefore also symmetric matrices) of a given rank are equivalent.

We now consider the case of a real vector space  $L$ . By Theorem 6.10, an arbitrary quadratic form  $\psi$  can again be written in the form

$$\psi(x) = \lambda_1 x_1^2 + \dots + \lambda_r x_r^2,$$

where all the  $\lambda_i$  are nonzero and  $r$  is the rank of the form  $\psi$ . But we cannot proceed so simply as in the complex case by setting  $e'_i = \sqrt{\lambda_i} e_i$ , since for  $\lambda_i < 0$ , the number  $\lambda_i$  does not have a real square root. Therefore, we must consider separately among the numbers  $\lambda_1, \dots, \lambda_r$ , those that are positive and those that are negative. Again changing the numeration of the vectors of the basis as necessary, we may assume that  $\lambda_1, \dots, \lambda_s$  are positive, and that  $\lambda_{s+1}, \dots, \lambda_r$  are negative. Now we can introduce a new basis by setting

$$e'_i = \sqrt{\lambda_i} e_i \quad \text{for } i \leq s, \quad e'_i = \sqrt{-\lambda_i} e_i \quad \text{for } i = s+1, \dots, r, \quad e'_i = e_i \quad \text{for } i > r.$$

In this basis, for a bilinear form  $\varphi$ , we have  $\varphi(\mathbf{e}'_i, \mathbf{e}'_j) = 0$  for  $i \neq j$ , and  $\varphi(\mathbf{e}'_i, \mathbf{e}'_i) = 1$  for  $i = 1, \dots, s$ ,  $\varphi(\mathbf{e}'_i, \mathbf{e}'_i) = -1$  for  $i = s + 1, \dots, r$ , and the quadratic form  $\psi$  will thus be brought into the form

$$\psi(\mathbf{x}) = x_1^2 + \dots + x_s^2 - x_{s+1}^2 - \dots - x_r^2. \quad (6.28)$$

Let us note one important special case.

**Definition 6.16** A real quadratic form  $\psi(\mathbf{x})$  is said to be *positive definite* if  $\psi(\mathbf{x}) > 0$  for every  $\mathbf{x} \neq \mathbf{0}$  and *negative definite* if  $\psi(\mathbf{x}) < 0$  for every  $\mathbf{x} \neq \mathbf{0}$ .

It is obvious that these notions are connected by a simple relationship: negative definite forms  $\psi(\mathbf{x})$  are equivalent to positive definite forms  $-\psi(\mathbf{x})$ , and conversely. Therefore, in the sequel, it will suffice to establish the basic properties of positive definite forms only, and the corresponding properties of negative definite forms will be obtained automatically.

Written in the form (6.28), a quadratic form on an  $n$ -dimensional vector space will be positive definite if  $s = n$ , and negative definite if  $s = 0$  and  $r = n$ .

The fundamental property of real quadratic forms is stated in the following theorem.

**Theorem 6.17** *For every basis in terms of which the real quadratic form  $\psi$  can be written in the form (6.28), the number  $s$  always has one and the same value.*

*Proof* Let us characterize  $s$  in a way that does not depend on reducing the quadratic form  $\psi$  to the form (6.28). Namely, let us prove that  $s$  is equal to the largest dimension among subspaces  $L' \subset L$  such that the restriction of  $\psi$  to  $L'$  is a positive definite quadratic form. To this end, we note first of all that for an arbitrary basis in which the form takes the form of (6.28), it is possible to find a subspace  $L'$  of dimension  $s$  on which the restriction of the form  $\psi$  gives a positive definite form. Namely, if the form  $\psi(\mathbf{x})$  is written in the form (6.28) in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , then we set  $L' = \langle \mathbf{e}_1, \dots, \mathbf{e}_s \rangle$ . It is obvious that the restriction of the form  $\psi$  to  $L'$  gives a positive definite quadratic form. Similarly, we may consider the set of vectors  $L''$  for which in the decomposition (6.28), the first  $s$  coordinates are equal to zero:  $x_1 = 0, \dots, x_s = 0$ . It is clear that this set is the vector subspace  $L'' = \langle \mathbf{e}_{s+1}, \mathbf{e}_{s+2}, \dots, \mathbf{e}_n \rangle$ , and for an arbitrary vector  $\mathbf{x} \in L''$ , we have the inequality  $\psi(\mathbf{x}) \leq 0$ .

Let us suppose that there exists a subspace  $M \subset L$  of dimension  $m > s$  such that the restriction of  $\psi$  to  $M$  gives a positive definite quadratic form. It is then obvious that  $\dim M + \dim L'' = m + n - s > n$ . By Corollary 3.42, the subspaces  $M$  and  $L''$  must have a common vector  $\mathbf{x} \neq \mathbf{0}$ . But since  $\mathbf{x} \in L''$ , it follows that  $\psi(\mathbf{x}) \leq 0$ , and since  $\mathbf{x} \in M$ , we have  $\psi(\mathbf{x}) > 0$ . This contradiction completes the proof of the theorem.  $\square$

**Definition 6.18** The number  $s$  from Theorem 6.17 that is the same no matter how a quadratic form is brought into the form (6.28) is called the *index of inertia* of the quadratic form  $\psi$ . In connection with this, Theorem 6.17 is often called the *law of inertia*.

Positive definite quadratic forms play an important role in the theory that we are expounding. By the theory developed thus far, to establish whether a quadratic form is positive definite, it is necessary to reduce it to canonical form and verify whether the relationship  $s = n$  is satisfied. However, there is a feature that makes it possible to determine positive definiteness from the matrix of the associated bilinear form written in an arbitrary basis. Suppose this matrix in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  has the form

$$\Phi = (\varphi_{ij}), \quad \text{where } \varphi_{ij} = \varphi(\mathbf{e}_i, \mathbf{e}_j).$$

The minor  $\Delta_i$  of the matrix  $\Phi$  at the intersection of the first  $i$  rows and first  $i$  columns is called a *leading principal minor*.

**Theorem 6.19** (Sylvester's criterion) *A quadratic form  $\psi$  is positive definite if and only if all leading principal minors of the matrix of the associated bilinear form are positive.*

*Proof* We shall show that if a quadratic form is positive definite, then all the  $\Delta_i$  are positive. We note as well that  $\Delta_n = |\Phi|$  is the determinant of the matrix of the form  $\varphi$ . In some basis, the form  $\psi$  is in canonical form, that is, its matrix in this basis has the form

$$\Phi' = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Since the quadratic form  $\psi$  is positive definite, it follows that all the  $\lambda_i$  are greater than 0, and clearly,  $|\Phi'| > 0$ . In view of formula (6.26) for replacing the matrix of a bilinear form by a change of basis along with the equality  $|C^*| = |C|$ , we obtain the relationship  $|\Phi'| = |\Phi| \cdot |C|^2$ , from which it follows that  $\Delta_n = |\Phi| > 0$ . Let us now consider the subspaces  $L_i = \langle \mathbf{e}_1, \dots, \mathbf{e}_i \rangle \subset L$  of dimension  $i \geq 1$ . The restriction of the quadratic form  $\psi(\mathbf{x})$  to  $L_i$  is clearly also a positive definite form. But the determinant of its matrix in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_i$  is equal to  $\Delta_i$ . Therefore,  $\Delta_i > 0$ , as we have shown.

Let us now show that conversely, from the condition  $\Delta_i > 0$  for all  $i = 1, \dots, n$ , it follows that the quadratic form  $\psi$  is positive definite. We shall prove this by induction on the dimension  $n$  of the space  $L$ .

It is clear that  $L_i \subset L$  for  $i = 1, \dots, n-1$ , and the leading principal minors  $\Delta_i$  in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the matrix of the form  $\psi$  restricted to the subspace  $L_i$  are the same as for the form  $\varphi$  in  $L$ . Therefore, the restriction of the quadratic form  $\psi$  to  $L_{n-1}$  may be assumed positive definite by the induction hypothesis. Consequently, the restriction  $\varphi(\mathbf{x}, \mathbf{y})$  to the subspace  $L_{n-1}$  is a nonsingular bilinear form, and so by Theorem 6.9, we have the decomposition  $L = L_{n-1} \oplus (L_{n-1})_{\varphi}^{\perp}$ , where  $\dim L_{n-1} = n-1$  and  $\dim (L_{n-1})_{\varphi}^{\perp} = 1$ . We may therefore express the vector  $\mathbf{e}_n$  in the form

$$\mathbf{e}_n = \mathbf{f}_n + \mathbf{y}, \quad \text{where } \mathbf{y} \in L_{n-1}, \mathbf{f}_n \in (L_{n-1})_{\varphi}^{\perp}. \quad (6.29)$$

We may represent an arbitrary vector  $\mathbf{x} \in \mathbb{L}$  as a linear combination of vectors of the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , that is, in the form  $\mathbf{x} = x_1\mathbf{e}_1 + \dots + x_{n-1}\mathbf{e}_{n-1} + x_n\mathbf{e}_n = \mathbf{u} + x_n\mathbf{e}_n$ , where  $\mathbf{u} \in \mathbb{L}_{n-1}$ . Substituting the expression (6.29) and setting  $\mathbf{u} + x_n\mathbf{y} = \mathbf{v}$ , we obtain

$$\mathbf{x} = \mathbf{v} + x_n\mathbf{f}_n, \quad \text{where } \mathbf{v} \in \mathbb{L}_{n-1}, \mathbf{f}_n \in (\mathbb{L}_{n-1})_{\varphi}^{\perp}. \quad (6.30)$$

This implies that the vectors  $\mathbf{v}$  and  $\mathbf{f}_n$  are orthogonal with respect to the bilinear form  $\varphi$ , that is,  $\varphi(\mathbf{v}, \mathbf{f}_n) = 0$ , and therefore, from the decomposition (6.30), we derive the equality

$$\psi(\mathbf{x}) = \psi(\mathbf{v}) + x_n^2\psi(\mathbf{f}_n). \quad (6.31)$$

We see, then, that in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_{n-1}, \mathbf{f}_n$ , the matrix of the bilinear form  $\varphi$  takes the form

$$\begin{pmatrix} \overline{\phantom{0}} & & 0 \\ \left| \begin{array}{c} \Phi' \end{array} \right| & & \vdots \\ \overline{\phantom{0}} & & 0 \\ 0 & \dots & 0 & \psi(\mathbf{f}_n) \end{pmatrix},$$

and for its determinant  $D_n$ , we obtain the expression  $D_n = |\Phi'| \cdot \psi(\mathbf{f}_n)$ . Since  $D_n > 0$  and  $|\Phi'| > 0$ , it then follows that  $\psi(\mathbf{f}_n) > 0$ . By the induction hypothesis, the term  $\psi(\mathbf{v})$  is positive in formula (6.31), and therefore,  $\psi(\mathbf{x}) > 0$  for every  $\mathbf{x} \neq \mathbf{0}$ .  $\square$

*Example 6.20* Sylvester's criterion has a beautiful application to the properties of algebraic equations. Consider a polynomial  $f(t)$  of degree  $n$  with real coefficients, about which we shall assume that its roots (real or complex)  $z_1, \dots, z_n$  are distinct. For each root  $z_k$ , we consider the linear form

$$l_k(\mathbf{x}) = x_1 + x_2z_k + \dots + x_nz_k^{n-1}, \quad (6.32)$$

and likewise the quadratic form

$$\psi(\mathbf{x}) = \sum_{k=1}^n l_k^2(x_1, \dots, x_n), \quad (6.33)$$

where  $\mathbf{x} = (x_1, \dots, x_n)$ .

Although among the roots  $z_k$  there may be some that are complex, the quadratic form (6.33) is always real. This is obvious for the terms  $l_k^2$  corresponding to the real roots  $z_k$ . Now, as regards the complex roots, it is well known that they come in complex conjugate pairs. Let  $z_k$  and  $z_j$  be complex conjugates of each other. Separating the coefficients  $l_k$  of the linear form into real and imaginary parts, we can write it in the form  $l_k = u_k + iv_k$ , where  $u_k$  and  $v_k$  are linear forms with real coefficients. Then  $l_j = u_k - iv_k$ , and for this pair of complex conjugate roots, we have the sum  $l_k^2 + l_j^2 = 2u_k^2 - 2v_k^2$ , which is a real quadratic form.

Thus the quadratic form (6.33) is real, and we have the following important criterion.

**Theorem 6.21** *All the roots of a polynomial  $f(t)$  are real if and only if the quadratic form (6.33) is positive definite.*

*Proof* If all the roots  $z_k$  are real, then all the linear forms  $l_k$  of (6.32) are real, and the sum on the right-hand side of (6.33) contains only nonnegative terms. It is clear that it is equal to zero only if  $l_k = 0$  for all  $k = 1, \dots, n$ . This condition gives us a system consisting of  $n$  linear homogeneous equations in  $n$  unknowns  $x_1, \dots, x_n$ . From formula (6.32), it is easy to see that the determinant of the matrix of this system is known to us already as a Vandermonde determinant; see formulas (2.32) and (2.33). It is different from zero, since all the roots  $z_k$  are distinct, and hence this system has only the null solution. This implies that  $\psi(\mathbf{x}) \geq 0$  and  $\psi(\mathbf{x}) = 0$  if and only if  $\mathbf{x} = \mathbf{0}$ , that is, the quadratic form (6.33) is positive definite.

Let us now prove the converse assertion. Let the quadratic form (6.33) be positive definite, and suppose the polynomial  $f(t)$  has  $r$  real roots and  $p$  pairs of complex roots, so that  $r + 2p = n$ . Then as we have seen,

$$\psi(\mathbf{x}) = \sum_{k=1}^p l_k^2 + 2 \sum_{j=1}^p (u_j^2 - v_j^2), \quad (6.34)$$

where the first sum extends over all real roots, and the second sum is over all pairs of complex conjugate roots.

Let us now show that if  $p > 0$ , then there exists a vector  $\mathbf{x} \neq \mathbf{0}$  such that

$$l_1(\mathbf{x}) = 0, \quad \dots, \quad l_r(\mathbf{x}) = 0, \quad u_1(\mathbf{x}) = 0, \quad \dots, \quad u_p(\mathbf{x}) = 0.$$

These equalities represent a system of  $r + p$  linear homogeneous equations in  $n$  unknowns  $x_1, \dots, x_n$ . Since the number of equations  $r + p$  is less than  $r + 2p = n$ , it follows that this system has a nontrivial solution,  $\mathbf{x} = (x_1, \dots, x_n)$ , for which the quadratic form (6.34) takes the form

$$\psi(\mathbf{x}) = -2 \sum_{j=1}^p v_j^2 \leq 0,$$

and moreover, the equality  $\psi(\mathbf{x}) = 0$  is possible only if  $v_j(\mathbf{x}) = 0$  for all  $j = 1, \dots, p$ . But then we obtain the equalities  $l_k(\mathbf{x}) = 0$  in general for all linear forms (6.32), which in view of the positive definiteness is possible only if  $\mathbf{x} = \mathbf{0}$ . We have thus obtained a contradiction to the fact that  $p > 0$ , that is, that the polynomial  $f(t)$  has at least one complex root.

The form (6.33) can be calculated explicitly, and then we can apply Sylvester's criterion to it. To this end, we observe that the coefficient of the monomial  $x_k^2$  on the right-hand side of (6.33) is equal to  $s_{2(k-1)} = z_1^{2(k-1)} + \dots + z_n^{2(k-1)}$ , while the



coefficient of the monomial  $x_i x_j$  (where  $i \neq j$ ) is equal to  $2s_{i+j-2} = 2(z_1^{i+j-2} + \dots + z_n^{i+j-2})$ . The sums  $s_k = \sum_{i=1}^n z_i^k$  are called *Newton sums*. It is known from the theory of symmetric functions that they can be expressed as polynomials in the coefficients of  $f(t)$ . Thus the matrix of a symmetric bilinear form associated with a quadratic form (6.33) has the form

$$\begin{pmatrix} s_0 & s_1 & \cdots & s_{n-1} \\ s_1 & s_2 & \cdots & s_n \\ \vdots & \vdots & \ddots & \vdots \\ s_{n-1} & s_n & \cdots & s_{2n-2} \end{pmatrix}.$$

Applying Sylvester's criterion to the form (6.33), we obtain the following result: all (distinct) roots of the polynomial  $f(t)$  are real if and only if the following inequality holds for all  $i = 1, \dots, n-1$ :

$$\begin{vmatrix} s_0 & s_1 & \cdots & s_{i-1} \\ s_1 & s_2 & \cdots & s_i \\ \vdots & \vdots & \ddots & \vdots \\ s_{i-1} & s_i & \cdots & s_{2i-2} \end{vmatrix} > 0. \quad \square$$

To illustrate this assertion, let us consider the simplest case,  $n = 2$ . Let  $f(t) = t^2 + pt + q$ . Then for the roots of the polynomial  $f(t)$  to be real and distinct is equivalent to the following two inequalities:

$$s_0 > 0, \quad \begin{vmatrix} s_0 & s_1 \\ s_1 & s_2 \end{vmatrix} > 0. \quad (6.35)$$

The first of these is satisfied for every polynomial, since  $s_0$  is simply its degree. If the roots of the polynomial  $f(t)$  are  $\alpha$  and  $\beta$ , then

$$s_0 = 2, \quad s_1 = \alpha + \beta = -p, \quad s_2 = \alpha^2 + \beta^2 = (\alpha + \beta)^2 - 2\alpha\beta = p^2 - 2q,$$

and inequality (6.35) yields  $2(p^2 - 2q) - p^2 = p^2 - 4q > 0$ . This is a criterion that one learns in secondary school: the roots of a quadratic trinomial are real and distinct if and only if the discriminant is positive.

We return now to complex vector spaces and consider certain functions in them that are more natural analogues of bilinear and quadratic forms than those examined at the beginning of this section.

**Definition 6.22** A function  $f(\mathbf{x})$  defined on a complex vector space  $L$  and taking complex values is said to be *semilinear* if it possesses the following properties:

$$\begin{aligned} f(\mathbf{x} + \mathbf{y}) &= f(\mathbf{x}) + f(\mathbf{y}), \\ f(\alpha \mathbf{x}) &= \bar{\alpha} f(\mathbf{x}), \end{aligned} \quad (6.36)$$

for arbitrary vectors  $\mathbf{x}$  and  $\mathbf{y}$  in the space  $L$  and complex scalar  $\alpha$  (here and below,  $\bar{\alpha}$  denotes the complex conjugate of  $\alpha$ ).

It is clear that for every choice of basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the space  $L$ , a semilinear function can be written in the form

$$f(\mathbf{x}) = \bar{x}_1 y_1 + \dots + \bar{x}_n y_n,$$

where the vector  $\mathbf{x}$  is equal to  $x_1 \mathbf{e}_1 + \dots + x_n \mathbf{e}_n$ , and the scalars  $y_i$  are equal to  $f(\mathbf{e}_i)$ .

**Definition 6.23** A function  $\varphi(\mathbf{x}, \mathbf{y})$  of two vectors in the complex vector space  $L$  is said to be *sesquilinear* if it is linear as a function of  $\mathbf{x}$  for fixed  $\mathbf{y}$  and semilinear as a function of  $\mathbf{y}$  for fixed  $\mathbf{x}$ .

The terminology “sesquilinear” indicates the “full” linearity of the first argument and semilinearity of the second. Semilinear and sesquilinear functions are also frequently called *forms*. In the sequel, we shall also use such a designation.

It is obvious that for an arbitrary choice of basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the space  $L$ , a sesquilinear form can be written in the form

$$\varphi(\mathbf{x}, \mathbf{y}) = \sum_{i,j=1}^n \varphi_{ij} x_i \bar{y}_j, \quad \text{where } \varphi_{ij} = \varphi(\mathbf{e}_i, \mathbf{e}_j), \quad (6.37)$$

and the vectors  $\mathbf{x}$  and  $\mathbf{y}$  are given by  $\mathbf{x} = x_1 \mathbf{e}_1 + \dots + x_n \mathbf{e}_n$  and  $\mathbf{y} = y_1 \mathbf{e}_1 + \dots + y_n \mathbf{e}_n$ . As in the case of a bilinear form, the matrix  $\Phi = (\varphi_{ij})$  with elements  $\varphi_{ij} = \varphi(\mathbf{e}_i, \mathbf{e}_j)$  as defined above is called the matrix of the sesquilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  in the chosen basis.

**Definition 6.24** A sesquilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  is said to be *Hermitian* if

$$\varphi(\mathbf{y}, \mathbf{x}) = \overline{\varphi(\mathbf{x}, \mathbf{y})} \quad (6.38)$$

for arbitrary choice of vectors  $\mathbf{x}$  and  $\mathbf{y}$ .

It is obvious that in the expression (6.37), the Hermitian nature of the form  $\varphi(\mathbf{x}, \mathbf{y})$  is expressed by the property  $\varphi_{ij} = \overline{\varphi_{ji}}$  of the coefficients  $\varphi_{ij}$  of its matrix  $\Phi$ , that is, by the relationship  $\Phi = \overline{\Phi}^*$ . A matrix exhibiting these properties is also called *Hermitian*.

After separating real and imaginary parts in  $\varphi(\mathbf{x}, \mathbf{y})$ , we obtain

$$\varphi(\mathbf{x}, \mathbf{y}) = u(\mathbf{x}, \mathbf{y}) + i v(\mathbf{x}, \mathbf{y}), \quad (6.39)$$

where  $u(\mathbf{x}, \mathbf{y})$  and  $v(\mathbf{x}, \mathbf{y})$  are functions of two vectors  $\mathbf{x}$  and  $\mathbf{y}$  of the complex space  $L$  taking real values. In the space  $L$ , multiplication by a real scalar is also defined, and so it may be viewed as a real vector space. We shall denote this real

vector space by  $L_{\mathbb{R}}$ . Obviously, in the space  $L_{\mathbb{R}}$ , the functions  $u(\mathbf{x}, \mathbf{y})$  and  $v(\mathbf{x}, \mathbf{y})$  are bilinear, and the property of the complex form  $\varphi(\mathbf{x}, \mathbf{y})$  being Hermitian implies that on  $L_{\mathbb{R}}$ , the bilinear form  $u(\mathbf{x}, \mathbf{y})$  is symmetric, while  $v(\mathbf{x}, \mathbf{y})$  is antisymmetric.

**Definition 6.25** A function  $\psi(\mathbf{x})$  on a complex vector space  $L$  is said to be *quadratic Hermitian* if it can be expressed in the form

$$\psi(\mathbf{x}) = \varphi(\mathbf{x}, \mathbf{x}) \quad (6.40)$$

for some Hermitian form  $\varphi(\mathbf{x}, \mathbf{y})$ .

From the definition of Hermitian form, it follows at once that the values of quadratic Hermitian functions are real.

**Theorem 6.26** A quadratic Hermitian function  $\psi(\mathbf{x})$  uniquely determines a Hermitian sesquilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  as presented in (6.40).

*Proof* By the definition of sesquilinearity, we have

$$\psi(\mathbf{x} + \mathbf{y}) = \psi(\mathbf{x}) + \psi(\mathbf{y}) + \varphi(\mathbf{x}, \mathbf{y}) + \overline{\varphi(\mathbf{x}, \mathbf{y})}. \quad (6.41)$$

Substituting here the expression (6.39), we obtain that

$$u(\mathbf{x}, \mathbf{y}) = \frac{1}{2}(\psi(\mathbf{x} + \mathbf{y}) - \psi(\mathbf{x}) - \psi(\mathbf{y})). \quad (6.42)$$

Similarly, from the relationship

$$\psi(\mathbf{x} + i\mathbf{y}) = \psi(\mathbf{x}) + \psi(i\mathbf{y}) + \varphi(\mathbf{x}, i\mathbf{y}) + \varphi(i\mathbf{y}, \mathbf{x}) \quad (6.43)$$

we obtain by the properties of being Hermitian and sesquilinearity that

$$\varphi(\mathbf{x}, i\mathbf{y}) = -i\varphi(\mathbf{x}, \mathbf{y}), \quad \varphi(i\mathbf{y}, \mathbf{x}) = \overline{\varphi(\mathbf{x}, i\mathbf{y})},$$

which yields

$$v(\mathbf{x}, \mathbf{y}) = \frac{1}{2}(\psi(\mathbf{x} + i\mathbf{y}) - \psi(\mathbf{x}) - \psi(i\mathbf{y})). \quad (6.44)$$

The expressions (6.42) and (6.44) thus obtained complete the proof of the theorem.  $\square$

**Theorem 6.27** A sesquilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  is Hermitian if and only if the function  $\psi(\mathbf{x})$  associated with it by relationship (6.40) assumes only real values.

*Proof* If a sesquilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  is Hermitian, then by definition (6.38), we have the equality  $\varphi(\mathbf{x}, \mathbf{x}) = \overline{\varphi(\mathbf{x}, \mathbf{x})}$  for all  $\mathbf{x} \in L$ , from which it follows that for an arbitrary vector  $\mathbf{x} \in L$ , the value  $\psi(\mathbf{x})$  is a real number.

On the other hand, if the values of the function  $\psi(\mathbf{x})$  are real, then arguing just as we did in the proof of Theorem 6.26, we obtain from formula (6.41), taking into account (6.38), that the value

$$\psi(\mathbf{x} + \mathbf{y}) - \psi(\mathbf{x}) - \psi(\mathbf{y}) = \varphi(\mathbf{x}, \mathbf{y}) + \varphi(\mathbf{y}, \mathbf{x})$$

is real. Substituting here the expression (6.39), we see that the sum  $v(\mathbf{x}, \mathbf{y}) + v(\mathbf{y}, \mathbf{x})$  is equal to zero, that is, the function  $v(\mathbf{x}, \mathbf{y})$  is antisymmetric.

Reasoning similarly, from formula (6.43), we conclude that the value

$$\psi(\mathbf{x} + i\mathbf{y}) - \psi(\mathbf{x}) - \psi(i\mathbf{y}) = \varphi(\mathbf{x}, i\mathbf{y}) + \varphi(i\mathbf{y}, \mathbf{x})$$

is also real. From the definition of semilinearity and sesquilinearity, we have the relationships  $\varphi(i\mathbf{y}, \mathbf{x}) = i\varphi(\mathbf{y}, \mathbf{x})$  and  $\varphi(\mathbf{x}, i\mathbf{y}) = -i\varphi(\mathbf{x}, \mathbf{y})$ . We thereby obtain that the number

$$i(\varphi(\mathbf{y}, \mathbf{x}) - \varphi(\mathbf{x}, \mathbf{y}))$$

is real, which by virtue of the expression (6.39) gives the equality  $u(\mathbf{y}, \mathbf{x}) - u(\mathbf{x}, \mathbf{y}) = 0$ ; that is, the function  $u(\mathbf{x}, \mathbf{y})$  is symmetric. Consequently, the form  $\varphi(\mathbf{x}, \mathbf{y})$  is Hermitian.  $\square$

Hermitian forms are the most natural complex analogues of symmetric forms. They exhibit analogous properties to those that we derived for symmetric forms in real vector spaces (with completely analogous proofs), namely reduction to canonical form, the law of inertia, the notion of positive definiteness, and Sylvester's criterion.

## Chapter 7

# Euclidean Spaces

The notions entering into the definition of a vector space do not provide a way of formulating multidimensional analogues of the length of a vector, the angle between vectors, and volumes. Yet such concepts appear in many branches of mathematics and physics, and we shall study such concepts in this chapter. All the vector spaces that we shall consider here will be real (with the exception of certain special cases in which complex vector spaces will be considered as a means of studying real spaces).

### 7.1 The Definition of a Euclidean Space

**Definition 7.1** A *Euclidean space* is a real vector space on which is defined a fixed symmetric bilinear form whose associated quadratic form is positive definite.

The vector space itself will be denoted as a rule by  $L$ , and the fixed symmetric bilinear form will be denoted by  $(\mathbf{x}, \mathbf{y})$ . Such an expression is also called the *inner product* of the vectors  $\mathbf{x}$  and  $\mathbf{y}$ . Let us now reformulate the definition of a Euclidean space using this terminology.

A *Euclidean space* is a real vector space  $L$  in which to every pair of vectors  $\mathbf{x}$  and  $\mathbf{y}$  there corresponds a real number  $(\mathbf{x}, \mathbf{y})$  such that the following conditions are satisfied:

- (1)  $(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}) = (\mathbf{x}_1, \mathbf{y}) + (\mathbf{x}_2, \mathbf{y})$  for all vectors  $\mathbf{x}_1, \mathbf{x}_2, \mathbf{y} \in L$ .
- (2)  $(\alpha \mathbf{x}, \mathbf{y}) = \alpha(\mathbf{x}, \mathbf{y})$  for all vectors  $\mathbf{x}, \mathbf{y} \in L$  and real number  $\alpha$ .
- (3)  $(\mathbf{x}, \mathbf{y}) = (\mathbf{y}, \mathbf{x})$  for all vectors  $\mathbf{x}, \mathbf{y} \in L$ .
- (4)  $(\mathbf{x}, \mathbf{x}) > 0$  for  $\mathbf{x} \neq \mathbf{0}$ .

Properties (1)–(3) show that the function  $(\mathbf{x}, \mathbf{y})$  is a symmetric bilinear form on  $L$ , and in particular, that  $(\mathbf{0}, \mathbf{y}) = 0$  for every vector  $\mathbf{y} \in L$ . It is only property (4) that expresses the specific character of a Euclidean space.

The expression  $(\mathbf{x}, \mathbf{x})$  is frequently denoted by  $(\mathbf{x}^2)$ ; it is called the *scalar square* of the vector  $\mathbf{x}$ . Thus property (4) implies that the quadratic form corresponding to the bilinear form  $(\mathbf{x}, \mathbf{y})$  is positive definite.

Let us point out some obvious consequences of these definitions. For a fixed vector  $\mathbf{y} \in L$ , where  $L$  is a Euclidean space, conditions (1) and (2) in the definition can be formulated in such a way that the function  $f_{\mathbf{y}}(\mathbf{x}) = (\mathbf{x}, \mathbf{y})$  with argument  $\mathbf{x}$  is linear. Thus we have a mapping  $\mathbf{y} \mapsto f_{\mathbf{y}}$  of the vector space  $L$  to  $L^*$ . Condition (4) in the definition of Euclidean space shows that the kernel of this mapping is equal to  $\{\mathbf{0}\}$ . Indeed,  $f_{\mathbf{y}} \neq \mathbf{0}$  for every  $\mathbf{y} \neq \mathbf{0}$ , since  $f_{\mathbf{y}}(\mathbf{y}) = (\mathbf{y}, \mathbf{y}) > 0$ . If the dimension of the space  $L$  is finite, then by Theorems 3.68 and 3.78, this mapping is an isomorphism. Moreover, we should note that in contrast to the construction used for proving Theorem 3.78, we have now constructed an isomorphism  $L \xrightarrow{\sim} L^*$  without using the specific choice of a basis in  $L$ . Thus we have a certain natural isomorphism  $L \xrightarrow{\sim} L^*$  defined only by the imposition of an inner product on  $L$ . In view of this, in the case of a finite-dimensional Euclidean space  $L$ , we shall in what follows sometimes identify  $L$  and  $L^*$ . In other words, as for any bilinear form, for the inner product  $(\mathbf{x}, \mathbf{y})$  there exists a unique linear transformation  $\mathcal{A} : L \rightarrow L^*$  such that  $(\mathbf{x}, \mathbf{y}) = \mathcal{A}(\mathbf{y})(\mathbf{x})$ . The previous reasoning shows that in the case of a Euclidean space, the transformation  $\mathcal{A}$  is an isomorphism, and in particular, the bilinear form  $(\mathbf{x}, \mathbf{y})$  is nonsingular. Let us give some examples of Euclidean spaces.

*Example 7.2* The plane, in which for  $(\mathbf{x}, \mathbf{y})$  is taken the well-known inner product of  $\mathbf{x}$  and  $\mathbf{y}$  as studied in analytic geometry, that is, the product of the vectors' lengths and the cosine of the angle between them, is a Euclidean space.

*Example 7.3* The space  $\mathbb{R}^n$  consisting of rows (or columns) of length  $n$ , in which the inner product of rows  $\mathbf{x} = (\alpha_1, \dots, \alpha_n)$  and  $\mathbf{y} = (\beta_1, \dots, \beta_n)$  is defined by the relation

$$(\mathbf{x}, \mathbf{y}) = \alpha_1\beta_1 + \alpha_2\beta_2 + \dots + \alpha_n\beta_n, \quad (7.1)$$

is a Euclidean space.

*Example 7.4* The vector space  $L$  consisting of polynomials of degree at most  $n$  with real coefficients, defined on some interval  $[a, b]$ , is a Euclidean space. For two polynomials  $f(t)$  and  $g(t)$ , their inner product is defined by the relation

$$(f, g) = \int_a^b f(t)g(t) dt. \quad (7.2)$$

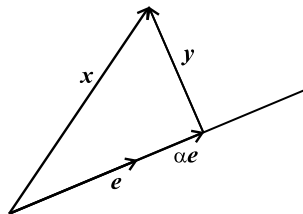
*Example 7.5* The vector space  $L$  consisting of all real-valued continuous functions on the interval  $[a, b]$  is a Euclidean space. For two such functions  $f(t)$  and  $g(t)$ , we shall define their inner product by equality (7.2).

Example 7.5 shows that a Euclidean space, like a vector space, does not have to be finite-dimensional.<sup>1</sup> In the sequel, we shall be concerned exclusively with finite-dimensional Euclidean spaces, on which the inner product is sometimes called the

---

<sup>1</sup>Infinite-dimensional Euclidean spaces are usually called *pre-Hilbert spaces*. An especially important role in a number of branches of mathematics and physics is played by the so-called *Hilbert*

**Fig. 7.1** Orthogonal projection



*scalar product* (because the inner product of two vectors is a scalar) or *dot product* (because the notation  $\mathbf{x} \cdot \mathbf{y}$  is frequently used instead of  $(\mathbf{x}, \mathbf{y})$ ).

*Example 7.6* Every subspace  $L'$  of a Euclidean space  $L$  is itself a Euclidean space if we define on it the form  $(\mathbf{x}, \mathbf{y})$  exactly as on the space  $L$ .

In analogy with Example 7.2, we make the following definition.

**Definition 7.7** The *length* of a vector  $\mathbf{x}$  in a Euclidean space is the nonnegative value  $\sqrt{(\mathbf{x}, \mathbf{x})}$ . The length of a vector  $\mathbf{x}$  is denoted by  $|\mathbf{x}|$ .

We note that we have here made essential use of property (4), by which the length of a nonnull vector is a positive number.

Following the same analogy, it is natural to define the *angle*  $\varphi$  between two vectors  $\mathbf{x}$  and  $\mathbf{y}$  by the condition

$$\cos \varphi = \frac{(\mathbf{x}, \mathbf{y})}{|\mathbf{x}| \cdot |\mathbf{y}|}, \quad 0 \leq \varphi \leq \pi. \quad (7.3)$$

However, such a number  $\varphi$  exists only if the expression on the right-hand side of equality (7.3) does not exceed 1 in absolute value. Such is indeed the case, and the proof of this fact will be our immediate objective.

**Lemma 7.8** Given a vector  $\mathbf{e} \neq \mathbf{0}$ , every vector  $\mathbf{x} \in L$  can be expressed in the form

$$\mathbf{x} = \alpha \mathbf{e} + \mathbf{y}, \quad (\mathbf{e}, \mathbf{y}) = 0, \quad (7.4)$$

for some scalar  $\alpha$  and vector  $\mathbf{y} \in L$ ; see Fig. 7.1.

*Proof* Setting  $\mathbf{y} = \mathbf{x} - \alpha \mathbf{e}$ , we obtain  $\alpha$  from the condition  $(\mathbf{e}, \mathbf{y}) = 0$ . This is equivalent to  $(\mathbf{x}, \mathbf{e}) = \alpha(\mathbf{e}, \mathbf{e})$ , which implies that  $\alpha = (\mathbf{x}, \mathbf{e})/|\mathbf{e}|^2$ . We remark that  $|\mathbf{e}| \neq 0$ , since by assumption,  $\mathbf{e} \neq \mathbf{0}$ .  $\square$

---

*spaces*, which are pre-Hilbert spaces that have the additional property of *completeness*, just for the case of infinite dimension. (Sometimes, in the definition of pre-Hilbert space, the condition  $(\mathbf{x}, \mathbf{x}) > 0$  is replaced by the weaker condition  $(\mathbf{x}, \mathbf{x}) \geq 0$ .)

**Definition 7.9** The vector  $\alpha \mathbf{e}$  from relation (7.4) is called the *orthogonal projection* of the vector  $\mathbf{x}$  onto the line  $\langle \mathbf{e} \rangle$ .

**Theorem 7.10** *The length of the orthogonal projection of a vector  $\mathbf{x}$  is at most its length  $|\mathbf{x}|$ .*

*Proof* Indeed, since by definition,  $\mathbf{x} = \alpha \mathbf{e} + \mathbf{y}$  and  $(\mathbf{e}, \mathbf{y}) = 0$ , it follows that

$$|\mathbf{x}|^2 = (\mathbf{x}^2) = (\alpha \mathbf{e} + \mathbf{y}, \alpha \mathbf{e} + \mathbf{y}) = |\alpha \mathbf{e}|^2 + |\mathbf{y}|^2 \geq |\alpha \mathbf{e}|^2,$$

and this implies that

$$|\mathbf{x}| \geq |\alpha \mathbf{e}|. \quad (7.5)$$

□

This leads directly to the following necessary theorem.

**Theorem 7.11** *For arbitrary vectors  $\mathbf{x}$  and  $\mathbf{y}$  in a Euclidean space, the following inequality holds:*

$$|(\mathbf{x}, \mathbf{y})| \leq |\mathbf{x}| \cdot |\mathbf{y}|. \quad (7.6)$$

*Proof* If one of the vectors  $\mathbf{x}$ ,  $\mathbf{y}$  is equal to zero, then the inequality (7.6) is obvious, and is reduced to the equality  $0 = 0$ . Now suppose that neither vector is the null vector. In this case, let us denote by  $\alpha \mathbf{y}$  the orthogonal projection of the vector  $\mathbf{x}$  onto the line  $\langle \mathbf{y} \rangle$ . Then by (7.4), we have the relationship  $\mathbf{x} = \alpha \mathbf{y} + \mathbf{z}$ , where  $(\mathbf{y}, \mathbf{z}) = 0$ . From this we obtain the equality

$$(\mathbf{x}, \mathbf{y}) = (\alpha \mathbf{y} + \mathbf{z}, \mathbf{y}) = (\alpha \mathbf{y}, \mathbf{y}) = \alpha |\mathbf{y}|^2.$$

This means that  $|(\mathbf{x}, \mathbf{y})| = |\alpha| \cdot |\mathbf{y}|^2 = |\alpha \mathbf{y}| \cdot |\mathbf{y}|$ . But by Theorem 7.10, we have the inequality  $|\alpha \mathbf{y}| \leq |\mathbf{x}|$ , and consequently,  $|(\mathbf{x}, \mathbf{y})| \leq |\mathbf{x}| \cdot |\mathbf{y}|$ . □

Inequality (7.6) goes by a number of names, but it is generally known as the Cauchy–Schwarz inequality. From it we can derive the well-known *triangle inequality* from elementary geometry. Indeed, suppose that the vectors  $\mathbf{x} = \overrightarrow{AB}$ ,  $\mathbf{y} = \overrightarrow{BC}$ ,  $\mathbf{z} = \overrightarrow{CA}$  correspond to the sides of a triangle  $ABC$ . Then we have the relationship  $\mathbf{x} + \mathbf{y} + \mathbf{z} = \mathbf{0}$ , from which with the help of (7.6) we obtain the inequality

$$\begin{aligned} |\mathbf{z}|^2 &= (\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) = |\mathbf{x}|^2 + 2(\mathbf{x}, \mathbf{y}) + |\mathbf{y}|^2 \leq |\mathbf{x}|^2 + 2|(\mathbf{x}, \mathbf{y})| + |\mathbf{y}|^2 \\ &\leq |\mathbf{x}|^2 + 2|\mathbf{x}| \cdot |\mathbf{y}| + |\mathbf{y}|^2 = (|\mathbf{x}| + |\mathbf{y}|)^2, \end{aligned}$$

from which clearly follows the triangle inequality  $|\mathbf{z}| \leq |\mathbf{x}| + |\mathbf{y}|$ .

Thus from Theorem 7.11 it follows that there exists a number  $\varphi$  that satisfies the equality (7.3). This number is what is called the *angle* between the vectors  $\mathbf{x}$  and  $\mathbf{y}$ . Condition (7.3) determines the angle uniquely if we assume that  $0 \leq \varphi \leq \pi$ .



**Definition 7.12** Two vectors  $\mathbf{x}$  and  $\mathbf{y}$  are said to be *orthogonal* if their inner product is equal to zero:  $(\mathbf{x}, \mathbf{y}) = 0$ .

Let us note that this repeats the definition given in Sect. 6.2 for a bilinear form  $\varphi(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{y})$ . By the definition given above in (7.3), the angle between orthogonal vectors is equal to  $\frac{\pi}{2}$ .

For a Euclidean space, there is a useful criterion for the linear independence of vectors. Let  $\mathbf{a}_1, \dots, \mathbf{a}_m$  be  $m$  vectors in the Euclidean space  $L$ .

**Definition 7.13** The *Gram determinant*, or *Gramian*, of a system of vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m$  is the determinant

$$G(\mathbf{a}_1, \dots, \mathbf{a}_m) = \begin{vmatrix} (\mathbf{a}_1, \mathbf{a}_1) & (\mathbf{a}_1, \mathbf{a}_2) & \cdots & (\mathbf{a}_1, \mathbf{a}_m) \\ (\mathbf{a}_2, \mathbf{a}_1) & (\mathbf{a}_2, \mathbf{a}_2) & \cdots & (\mathbf{a}_2, \mathbf{a}_m) \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{a}_m, \mathbf{a}_1) & (\mathbf{a}_m, \mathbf{a}_2) & \cdots & (\mathbf{a}_m, \mathbf{a}_m) \end{vmatrix}. \quad (7.7)$$

**Theorem 7.14** If the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m$  are linearly dependent, then the Gram determinant  $G(\mathbf{a}_1, \dots, \mathbf{a}_m)$  is equal to zero, while if they are linearly independent, then  $G(\mathbf{a}_1, \dots, \mathbf{a}_m) > 0$ .

*Proof* If the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m$  are linearly dependent, then as was shown in Sect. 3.2, one of the vectors can be expressed as a linear combination of the others. Let it be the vector  $\mathbf{a}_m$ , that is,  $\mathbf{a}_m = \alpha_1 \mathbf{a}_1 + \cdots + \alpha_{m-1} \mathbf{a}_{m-1}$ . Then from the properties of the inner product, it follows that for every  $i = 1, \dots, m$ , we have the equality

$$(\mathbf{a}_m, \mathbf{a}_i) = \alpha_1 (\mathbf{a}_1, \mathbf{a}_i) + \alpha_2 (\mathbf{a}_2, \mathbf{a}_i) + \cdots + \alpha_{m-1} (\mathbf{a}_{m-1}, \mathbf{a}_i).$$

From this it is clear that if we subtract from the last row of the determinant (7.7), all the previous rows multiplied by coefficients  $\alpha_1, \dots, \alpha_{m-1}$ , then we obtain a determinant with a row consisting entirely of zeros. Therefore,  $G(\mathbf{a}_1, \dots, \mathbf{a}_m) = 0$ .

Now suppose that vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m$  are linearly independent. Let us consider in the subspace  $L' = \langle \mathbf{a}_1, \dots, \mathbf{a}_m \rangle$ , the quadratic form  $(\mathbf{x}^2)$ . Setting  $\mathbf{x} = \alpha_1 \mathbf{a}_1 + \cdots + \alpha_m \mathbf{a}_m$ , we may write it in the form

$$((\alpha_1 \mathbf{a}_1 + \cdots + \alpha_m \mathbf{a}_m)^2) = \sum_{i,j=1}^m \alpha_i \alpha_j (\mathbf{a}_i, \mathbf{a}_j).$$

It is easily seen that this quadratic form is positive definite, and its determinant coincides with the Gram determinant  $G(\mathbf{a}_1, \dots, \mathbf{a}_m)$ . By Theorem 6.19, it now follows that  $G(\mathbf{a}_1, \dots, \mathbf{a}_m) > 0$ .  $\square$

Theorem 7.14 is a broad generalization of the Cauchy–Schwarz inequality. Indeed, for  $m = 2$ , inequality (7.6) is obvious (it becomes an equality) if vectors  $\mathbf{x}$

and  $\mathbf{y}$  are linearly dependent. However, if  $\mathbf{x}$  and  $\mathbf{y}$  are linearly independent, then their Gram determinant is equal to

$$G(\mathbf{x}, \mathbf{y}) = \begin{vmatrix} (\mathbf{x}, \mathbf{x}) & (\mathbf{x}, \mathbf{y}) \\ (\mathbf{x}, \mathbf{y}) & (\mathbf{y}, \mathbf{y}) \end{vmatrix}.$$

The inequality  $G(\mathbf{x}, \mathbf{y}) > 0$  established in Theorem 7.14 gives us (7.6). In particular, we see that inequality (7.6) becomes an equality *only* if the vectors  $\mathbf{x}$  and  $\mathbf{y}$  are proportional. We remark that this is easy to derive if we examine the proof of Theorem 7.11.

**Definition 7.15** Vectors  $\mathbf{e}_1, \dots, \mathbf{e}_m$  in a Euclidean space form an *orthonormal system* if

$$(\mathbf{e}_i, \mathbf{e}_j) = 0 \quad \text{for } i \neq j, \quad (\mathbf{e}_i, \mathbf{e}_i) = 1, \quad (7.8)$$

that is, if these vectors are mutually orthogonal and the length of each of them is equal to 1. If  $m = n$  and the vectors  $\mathbf{e}_1, \dots, \mathbf{e}_n$  form a basis of the space, then such a basis is called an *orthonormal basis*.

It is obvious that the Gram determinant of an orthonormal basis is equal to 1.

We shall now use the fact that a quadratic form  $(\mathbf{x}^2)$  is positive definite and apply to it formula (6.28), in which by the definition of positive definiteness,  $s = n$ . This result can now be reformulated as an assertion about the existence of a basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the space  $L$  in which the scalar square of a vector  $\mathbf{x} = \alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n$  is equal to the sum of the squares of its coordinates, that is,  $(\mathbf{x}^2) = \alpha_1^2 + \dots + \alpha_n^2$ . In other words, we have the following result.

**Theorem 7.16** *Every Euclidean space has an orthonormal basis.*

**Remark 7.17** In an orthonormal basis, the inner product of  $\mathbf{x} = (\alpha_1, \dots, \alpha_n)$  and  $\mathbf{y} = (\beta_1, \dots, \beta_n)$  has a particularly simple form, given by formula (7.1). Accordingly, in an orthonormal basis, the scalar square of an arbitrary vector is equal to the sum of the squares of its coordinates, while its length is equal to the square root of the sum of the squares.

The lemma establishing the decomposition (7.4) has an important and far-reaching generalization. To formulate it, we recall that in Sect. 3.7, for every subspace  $L' \subset L$  we defined its annihilator  $(L')^a \subset L^*$ , while earlier in this section, we showed that an arbitrary Euclidean space  $L$  of finite dimension can be identified with its dual space  $L^*$ . As a result, we can view  $(L')^a$  as a subspace of the original space  $L$ . In this light, we shall call it the *orthogonal complement* of the subspace  $L'$  and denote it by  $(L')^\perp$ . If we recall the relevant definitions, we obtain that the orthogonal complement  $(L')^\perp$  of the subspace  $L' \subset L$  consists of all vectors  $\mathbf{y} \in L$  for which the following condition holds:

$$(\mathbf{x}, \mathbf{y}) = 0 \quad \text{for all } \mathbf{x} \in L'. \quad (7.9)$$

On the other hand,  $(L')^\perp$  is the subspace  $(L')^\perp_\varphi$ , defined for the case that the bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  is given by  $\varphi(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{y})$ ; see p. 198.

A basic property of the orthogonal complement in a finite-dimensional Euclidean space is contained in the following theorem.

**Theorem 7.18** *For an arbitrary subspace  $L_1$  of a finite-dimensional Euclidean space  $L$ , the following holds:*

$$L = L_1 \oplus L_1^\perp. \quad (7.10)$$

In the case  $L_1 = \langle \mathbf{e} \rangle$ , Theorem 7.18 follows from Lemma 7.8.

*Proof of Theorem 7.18* In the previous chapter, we saw that every quadratic form  $\psi(\mathbf{x})$  in some basis of a vector space  $L$  can be reduced to the canonical form (6.22), and in the case of a real vector space, to the form (6.28) for some scalars  $0 \leq s \leq r$ , where  $s$  is the index of inertia and  $r$  is the rank of the quadratic form  $\psi(\mathbf{x})$ , or equivalently, the rank of the symmetric bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  associated with  $\psi(\mathbf{x})$  by the relationship (6.11). We recall that a bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  is nonsingular if  $r = n$ , where  $n = \dim L$ .

The condition of positive definiteness for the form  $\psi(\mathbf{x})$  is equivalent to the condition that all scalars  $\lambda_1, \dots, \lambda_n$  in (6.22) be positive, or equivalently, that the equality  $s = r = n$  hold in formula (6.28). From this it follows that a symmetric bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  associated with a positive definite quadratic form  $\psi(\mathbf{x})$  is nonsingular on the space  $L$  as well as on every subspace  $L' \subset L$ . To complete the proof, it suffices to recall that by definition, the quadratic form  $(\mathbf{x}^2)$  associated with the inner product  $(\mathbf{x}, \mathbf{y})$  is positive definite and to use Theorem 6.9 for the bilinear form  $\varphi(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{y})$ .  $\square$

From relationship (3.54) for the annihilator (see Sect. 3.7) or from Theorem 7.18, it follows that

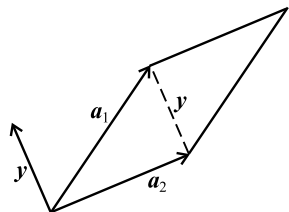
$$\dim(L_1)^\perp = \dim L - \dim L_1.$$

The map that is the projection of the space  $L$  onto the subspace  $L_1$  parallel to  $L_1^\perp$  (see the definition on p. 103) is called the *orthogonal projection* of  $L$  onto  $L_1$ . Then the projection of the vector  $\mathbf{x} \in L$  onto the subspace  $L_1$  is called its *orthogonal projection* onto  $L_1$ . This is a natural generalization of the notion introduced above of orthogonal projection of a vector onto a line. Similarly, for an arbitrary subset  $X \subset L$ , we can define its orthogonal projection onto  $L_1$ .

The Gram determinant is connected to the notion of *volume* in a Euclidean space, generalizing the notion of the length of a vector.

**Definition 7.19** The *parallelepiped spanned by vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m$*  is the collection of all vectors  $\alpha_1 \mathbf{a}_1 + \dots + \alpha_m \mathbf{a}_m$  for all  $0 \leq \alpha_i \leq 1$ . It is denoted by  $\Pi(\mathbf{a}_1, \dots, \mathbf{a}_m)$ . A *base* of the parallelepiped  $\Pi(\mathbf{a}_1, \dots, \mathbf{a}_m)$  is a parallelepiped spanned by any  $m - 1$  vectors among  $\mathbf{a}_1, \dots, \mathbf{a}_m$ , for example,  $\Pi(\mathbf{a}_1, \dots, \mathbf{a}_{m-1})$ .

**Fig. 7.2** Altitude of a parallelepiped



In the case of the plane (see Example 7.2), we have parallelepipeds  $\Pi(\mathbf{a}_1)$  and  $\Pi(\mathbf{a}_1, \mathbf{a}_2)$ . By definition,  $\Pi(\mathbf{a}_1)$  is the segment whose beginning and end coincide with the beginning and end of the vector  $\mathbf{a}_1$ , while  $\Pi(\mathbf{a}_1, \mathbf{a}_2)$  is the parallelogram constructed from the vectors  $\mathbf{a}_1$  and  $\mathbf{a}_2$ .

We return now to the consideration of an arbitrary parallelepiped

$$\Pi(\mathbf{a}_1, \dots, \mathbf{a}_m),$$

and we define the subspace  $L_1 = \langle \mathbf{a}_1, \dots, \mathbf{a}_{m-1} \rangle$ . To this case we may apply the notion introduced above of orthogonal projection of the space  $L$ . By the decomposition (7.10), the vector  $\mathbf{a}_m$  can be uniquely represented in the form  $\mathbf{a}_m = \mathbf{x} + \mathbf{y}$ , where  $\mathbf{x} \in L_1$  and  $\mathbf{y} \in L_1^\perp$ . The vector  $\mathbf{y}$  is called the *altitude* of the parallelepiped  $\Pi(\mathbf{a}_1, \dots, \mathbf{a}_m)$  dropped to the base  $\Pi(\mathbf{a}_1, \dots, \mathbf{a}_{m-1})$ . The construction we have described is depicted in Fig. 7.2 for the case of the plane.

Now we can introduce the concept of *volume* of a parallelepiped

$$\Pi(\mathbf{a}_1, \dots, \mathbf{a}_m),$$

or more precisely, its *unoriented volume*. This is by definition a nonnegative number, denoted by  $V(\mathbf{a}_1, \dots, \mathbf{a}_m)$  and defined by induction on  $m$ . In the case  $m = 1$ , it is equal to  $V(\mathbf{a}_1) = |\mathbf{a}_1|$ , and in the general case,  $V(\mathbf{a}_1, \dots, \mathbf{a}_m)$  is the product of  $V(\mathbf{a}_1, \dots, \mathbf{a}_{m-1})$  and the length of the altitude of the parallelepiped  $\Pi(\mathbf{a}_1, \dots, \mathbf{a}_m)$  dropped to the base  $\Pi(\mathbf{a}_1, \dots, \mathbf{a}_{m-1})$ .

The following is a numerical expression for the unoriented volume:

$$V^2(\mathbf{a}_1, \dots, \mathbf{a}_m) = G(\mathbf{a}_1, \dots, \mathbf{a}_m). \quad (7.11)$$

This relationship shows the geometric meaning of the Gram determinant.

Formula (7.11) is obvious for  $m = 1$ , and in the general case, it is proved by induction on  $m$ . According to (7.10), we may represent the vector  $\mathbf{a}_m$  in the form  $\mathbf{a}_m = \mathbf{x} + \mathbf{y}$ , where  $\mathbf{x} \in L_1 = \langle \mathbf{a}_1, \dots, \mathbf{a}_{m-1} \rangle$  and  $\mathbf{y} \in L_1^\perp$ . Then  $\mathbf{a}_m = \alpha_1 \mathbf{a}_1 + \dots + \alpha_{m-1} \mathbf{a}_{m-1} + \mathbf{y}$ . We note that  $\mathbf{y}$  is the altitude of our parallelepiped dropped to the base  $\Pi(\mathbf{a}_1, \dots, \mathbf{a}_{m-1})$ . Let us recall formula (7.7) for the Gram determinant and subtract from its last column, each of the other columns multiplied by  $\alpha_1, \dots, \alpha_{m-1}$ .

As a result, we obtain

$$G(\mathbf{a}_1, \dots, \mathbf{a}_m) = \begin{vmatrix} (\mathbf{a}_1, \mathbf{a}_1) & (\mathbf{a}_1, \mathbf{a}_2) & \cdots & 0 \\ (\mathbf{a}_2, \mathbf{a}_1) & (\mathbf{a}_2, \mathbf{a}_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{a}_{m-1}, \mathbf{a}_1) & (\mathbf{a}_{m-1}, \mathbf{a}_2) & \cdots & 0 \\ (\mathbf{a}_m, \mathbf{a}_1) & (\mathbf{a}_m, \mathbf{a}_2) & \cdots & (\mathbf{y}, \mathbf{a}_m) \end{vmatrix}, \quad (7.12)$$

and moreover,  $(\mathbf{y}, \mathbf{a}_m) = (\mathbf{y}, \mathbf{y}) = |\mathbf{y}|^2$ , since  $\mathbf{y} \in L_1^\perp$ .

Expanding the determinant (7.12) along its last column, we obtain the equality

$$G(\mathbf{a}_1, \dots, \mathbf{a}_m) = G(\mathbf{a}_1, \dots, \mathbf{a}_{m-1})|\mathbf{y}|^2.$$

Let us recall that by construction,  $\mathbf{y}$  is the altitude of the parallelepiped  $\Pi(\mathbf{a}_1, \dots, \mathbf{a}_m)$  dropped to the base  $\Pi(\mathbf{a}_1, \dots, \mathbf{a}_{m-1})$ . By the induction hypothesis, we have  $G(\mathbf{a}_1, \dots, \mathbf{a}_{m-1}) = V^2(\mathbf{a}_1, \dots, \mathbf{a}_{m-1})$ , and this implies

$$G(\mathbf{a}_1, \dots, \mathbf{a}_m) = V^2(\mathbf{a}_1, \dots, \mathbf{a}_{m-1})|\mathbf{y}|^2 = V^2(\mathbf{a}_1, \dots, \mathbf{a}_m).$$

Thus the concept of unoriented volume that we have introduced differs from the volume and area about which we spoke in Sects. 2.1 and 2.6, since the unoriented volume cannot assume negative values. This explains the term “unoriented.” We shall now formulate a second way of looking at the volume of a parallelepiped, one that generalizes the notions of volume and area about which we spoke earlier and differs from unoriented volume by the sign  $\pm 1$ . By Theorem 7.14, of interest is only the case in which the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m$  are linearly independent. Then we may consider the space  $L = \langle \mathbf{a}_1, \dots, \mathbf{a}_m \rangle$  with basis  $\mathbf{a}_1, \dots, \mathbf{a}_m$ .

Thus we are given  $n$  vectors  $\mathbf{a}_1, \dots, \mathbf{a}_n$ , where  $n = \dim L$ . We consider the matrix  $A$ , whose  $j$ th column consists of the coordinates of the vector  $\mathbf{a}_j$  relative to some orthonormal basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ :

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

An easy verification shows that in the matrix  $A^*A$ , the intersection of the  $i$ th row and  $j$ th column contains the element  $(\mathbf{a}_i, \mathbf{a}_j)$ . This implies that the determinant of the matrix  $A^*A$  is equal to  $G(\mathbf{a}_1, \dots, \mathbf{a}_n)$ , and in view of the equalities  $|A^*A| = |A^*| \cdot |A| = |A|^2$ , we obtain  $|A|^2 = G(\mathbf{a}_1, \dots, \mathbf{a}_n)$ . On the other hand, from formula (7.11), it follows that  $G(\mathbf{a}_1, \dots, \mathbf{a}_n) = V^2(\mathbf{a}_1, \dots, \mathbf{a}_n)$ , and this implies that

$$|A| = \pm V(\mathbf{a}_1, \dots, \mathbf{a}_n).$$

The determinant of the matrix  $A$  is called the *oriented volume* of the  $n$ -dimensional parallelepiped  $\Pi(\mathbf{a}_1, \dots, \mathbf{a}_n)$ . It is denoted by  $v(\mathbf{a}_1, \dots, \mathbf{a}_n)$ . Thus the oriented and

unoriented volumes are related by the equality

$$V(\mathbf{a}_1, \dots, \mathbf{a}_n) = |v(\mathbf{a}_1, \dots, \mathbf{a}_n)|.$$

Since the determinant of a matrix does not change under the transpose operation, it follows that  $v(\mathbf{a}_1, \dots, \mathbf{a}_n) = |A^*|$ . In other words, for computing the oriented volume, one may write the coordinates of the generators of the parallelepiped  $\mathbf{a}_i$  not in the columns of the matrix, but in the rows, which is sometimes more convenient.

It is obvious that the sign of the oriented volume *depends* on the choice of orthonormal basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ . This dependence is suggested by the term “oriented.” We shall have more to say about this in Sect. 7.3.

The volume possesses some important properties.

**Theorem 7.20** *Let  $\mathcal{C} : \mathbb{L} \rightarrow \mathbb{L}$  be a linear transformation of the Euclidean space  $\mathbb{L}$  of dimension  $n$ . Then for any  $n$  vectors  $\mathbf{a}_1, \dots, \mathbf{a}_n$  in this space, one has the relationship*

$$v(\mathcal{C}(\mathbf{a}_1), \dots, \mathcal{C}(\mathbf{a}_n)) = |\mathcal{C}| v(\mathbf{a}_1, \dots, \mathbf{a}_n). \quad (7.13)$$

*Proof* We shall choose an orthonormal basis of the space  $\mathbb{L}$ . Suppose that the transformation  $\mathcal{C}$  has matrix  $C$  in this basis and that the coordinates  $\alpha_1, \dots, \alpha_n$  of an arbitrary vector  $\mathbf{a}$  are related to the coordinates  $\beta_1, \dots, \beta_n$  of its image  $\mathcal{C}(\mathbf{a})$  by the relationship (3.25), or in matrix notation, (3.27). Let  $A$  be the matrix whose columns consist of the coordinates of the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_n$ , and let  $A'$  be the matrix whose columns consist of the coordinates of the vectors  $\mathcal{C}(\mathbf{a}_1), \dots, \mathcal{C}(\mathbf{a}_n)$ . Then it is obvious that we have the relationship  $A' = CA$ , from which it follows that  $|A'| = |C| \cdot |A|$ .

To complete the proof, it remains to note that  $|\mathcal{C}| = |C|$ , and by the definition of oriented volume, we have the equalities  $v(\mathbf{a}_1, \dots, \mathbf{a}_n) = |A|$  and  $v(\mathcal{C}(\mathbf{a}_1), \dots, \mathcal{C}(\mathbf{a}_n)) = |A'|$ .  $\square$

It follows from this theorem, of course, that

$$V(\mathcal{C}(\mathbf{a}_1), \dots, \mathcal{C}(\mathbf{a}_n)) = ||A|| V(\mathbf{a}_1, \dots, \mathbf{a}_n), \quad (7.14)$$

where  $||A||$  denotes the absolute value of the determinant of the matrix  $A$ .

Using the concepts introduced thus far, we may define an analogue of the volume  $V(M)$  for a very broad class of sets  $M$  containing all the sets actually encountered in mathematics and physics. This is the subject of what is called *measure theory*, but since it is a topic that is rather far removed from linear algebra, it will not concern us here. Let us note only that the important relationship (7.14) remains valid here:

$$V(\mathcal{C}(M)) = ||A|| V(M). \quad (7.15)$$

An interesting example of a set in an  $n$ -dimensional Euclidean space is the *ball*  $B(r)$  of radius  $r$ , namely the set of all vectors  $\mathbf{x} \in \mathbb{L}$  such that  $|\mathbf{x}| \leq r$ . The set of vectors  $\mathbf{x} \in \mathbb{L}$  for which  $|\mathbf{x}| = r$  is called the *sphere*  $S(r)$  of radius  $r$ . From the relationship (7.15) it follows that  $V(B(r)) = V_n r^n$ , where  $V_n = V(B(1))$ . The calculation of the

interesting geometric constant  $V_n$  is a question from analysis, related to the theory of the *gamma function*  $\Gamma$ . Here we shall simply quote the result:

$$V_n = \frac{\pi^{n/2}}{\Gamma(n/2 + 1)}.$$

It follows from the theory of the gamma function that if  $n$  is an even number ( $n = 2m$ ), then  $V_n = \pi^m/m!$ , and if  $n$  is odd ( $n = 2m + 1$ ), then  $V_n = 2^{m+1}\pi^m/(1 \cdot 3 \cdots (2m + 1))$ .

## 7.2 Orthogonal Transformations

Let  $L_1$  and  $L_2$  be Euclidean spaces of the same dimension with inner products  $(\mathbf{x}, \mathbf{y})_1$  and  $(\mathbf{x}, \mathbf{y})_2$  defined on them. We shall denote the length of a vector  $\mathbf{x}$  in the spaces  $L_1$  and  $L_2$  by  $|\mathbf{x}|_1$  and  $|\mathbf{x}|_2$ , respectively.

**Definition 7.21** An *isomorphism* of Euclidean spaces  $L_1$  and  $L_2$  is an isomorphism  $\mathcal{A} : L_1 \rightarrow L_2$  of the underlying vector spaces that preserves the inner product, that is, for arbitrary vectors  $\mathbf{x}, \mathbf{y} \in L_1$ , the following relationship holds:

$$(\mathbf{x}, \mathbf{y})_1 = (\mathcal{A}(\mathbf{x}), \mathcal{A}(\mathbf{y}))_2. \quad (7.16)$$

If we substitute the vector  $\mathbf{y} = \mathbf{x}$  into equality (7.16), we obtain that  $|\mathbf{x}|_1^2 = |\mathcal{A}(\mathbf{x})|_2^2$ , and this implies that  $|\mathbf{x}|_1 = |\mathcal{A}(\mathbf{x})|_2$ , that is, the isomorphism  $\mathcal{A}$  preserves the lengths of vectors.

Conversely, if  $\mathcal{A} : L_1 \rightarrow L_2$  is an isomorphism of vector spaces that preserves the lengths of vectors, then  $|\mathcal{A}(\mathbf{x} + \mathbf{y})|_2^2 = |\mathbf{x} + \mathbf{y}|_1^2$ , and therefore,

$$|\mathcal{A}(\mathbf{x})|_2^2 + 2(\mathcal{A}(\mathbf{x}), \mathcal{A}(\mathbf{y}))_2 + |\mathcal{A}(\mathbf{y})|_2^2 = |\mathbf{x}|_1^2 + 2(\mathbf{x}, \mathbf{y})_1 + |\mathbf{y}|_1^2.$$

But by assumption, we also have the equalities  $|\mathcal{A}(\mathbf{x})|_2 = |\mathbf{x}|_1$  and  $|\mathcal{A}(\mathbf{y})|_2 = |\mathbf{y}|_1$ , which implies that  $(\mathbf{x}, \mathbf{y})_1 = (\mathcal{A}(\mathbf{x}), \mathcal{A}(\mathbf{y}))_2$ . This, strictly speaking, is a consequence of the fact (Theorem 6.6) that a symmetric bilinear form  $(\mathbf{x}, \mathbf{y})$  is determined by the quadratic form  $(\mathbf{x}, \mathbf{x})$ , and here we have simply repeated the proof given in Sect. 4.1.

If the spaces  $L_1$  and  $L_2$  have the same dimension, then from the fact that the linear transformation  $\mathcal{A} : L_1 \rightarrow L_2$  preserves the lengths of vectors, it already follows that it is an isomorphism. Indeed, as we saw in Sect. 3.5, it suffices to verify that the kernel of the transformation  $\mathcal{A}$  is equal to  $(\mathbf{0})$ . But if  $\mathcal{A}(\mathbf{x}) = \mathbf{0}$ , then  $|\mathcal{A}(\mathbf{x})|_2 = 0$ , which implies that  $|\mathbf{x}|_1 = 0$ , that is,  $\mathbf{x} = \mathbf{0}$ .

**Theorem 7.22** All Euclidean spaces of a given finite dimension are isomorphic to each other.

*Proof* From the existence of an orthonormal basis, it follows at once that every  $n$ -dimensional Euclidean space is isomorphic to the Euclidean space in Example 7.3. Indeed, let  $e_1, \dots, e_n$  be an orthonormal basis of a Euclidean space  $L$ . Assigning to each vector  $x \in L$  the row of its coordinates in the basis  $e_1, \dots, e_n$ , we obtain an isomorphism of the space  $L$  and the space  $\mathbb{R}^n$  of rows of length  $n$  with inner product (7.1) (see the remarks on p. 218). It is easily seen that isomorphism is an equivalence relation (p. xii) on the set of Euclidean spaces, and by transitivity, it follows that all Euclidean spaces of dimension  $n$  are isomorphic to each other.  $\square$

Theorem 7.22 is analogous to Theorem 3.64 for vector spaces, and its general meaning is the same (this is elucidated in detail in Sect. 3.5). For example, using Theorem 7.22, we could have proved the inequality (7.6) differently from how it was done in the preceding section. Indeed, it is completely obvious (the inequality is reduced to an equality) if the vectors  $x$  and  $y$  are linearly dependent. If, on the other hand, they are linearly independent, then we can consider the subspace  $L' = \langle x, y \rangle$ . By Theorem 7.22, it is isomorphic to the plane (Example 7.2 in the previous section), where this inequality is well known. Therefore, it must also be correct for arbitrary vectors  $x$  and  $y$ .

**Definition 7.23** A linear transformation  $\mathcal{U}$  of a Euclidean space  $L$  into itself that preserves the inner product, that is, satisfies the condition that for all vectors  $x$  and  $y$ ,

$$(x, y) = (\mathcal{U}(x), \mathcal{U}(y)), \quad (7.17)$$

is said to be *orthogonal*.

This is clearly a special case of an isomorphism of Euclidean spaces  $L_1$  and  $L_2$  that coincide.

It is also easily seen that an orthogonal transformation  $\mathcal{U}$  takes an orthonormal basis to another orthonormal basis, since from the conditions (7.8) and (7.17), it follows that  $\mathcal{U}(e_1), \dots, \mathcal{U}(e_n)$  is an orthonormal basis if  $e_1, \dots, e_n$  is. Conversely, if a linear transformation  $\mathcal{U}$  takes *some* orthonormal basis  $e_1, \dots, e_n$  to another orthonormal basis, then for vectors  $x = \alpha_1 e_1 + \dots + \alpha_n e_n$  and  $y = \beta_1 e_1 + \dots + \beta_n e_n$ , we have

$$\mathcal{U}(x) = \alpha_1 \mathcal{U}(e_1) + \dots + \alpha_n \mathcal{U}(e_n), \quad \mathcal{U}(y) = \beta_1 \mathcal{U}(e_1) + \dots + \beta_n \mathcal{U}(e_n).$$

Since both  $e_1, \dots, e_n$  and  $\mathcal{U}(e_1), \dots, \mathcal{U}(e_n)$  are orthonormal bases, it follows by (7.1) that both the left- and right-hand sides of relationship (7.17) are equal to the expression  $\alpha_1 \beta_1 + \dots + \alpha_n \beta_n$ , that is, relationship (7.17) is satisfied, and this implies that  $\mathcal{U}$  is an orthogonal transformation.

We note the following important reformulation of this fact: for any two orthonormal bases of a Euclidean space, there exists a unique orthogonal transformation that takes the first basis into the second.

Let  $U = (u_{ij})$  be the matrix of a linear transformation  $\mathcal{U}$  in some orthonormal basis  $e_1, \dots, e_n$ . It follows from what has gone before that the transformation  $\mathcal{U}$  is



orthogonal if and only if the vectors  $\mathcal{U}(\mathbf{e}_1), \dots, \mathcal{U}(\mathbf{e}_n)$  form an orthonormal basis. But by the definition of the matrix  $U$ , the vector  $\mathcal{U}(\mathbf{e}_i)$  is equal to  $\sum_{k=1}^n u_{ki} \mathbf{e}_k$ , and since  $\mathbf{e}_1, \dots, \mathbf{e}_n$  is an orthonormal basis, we have

$$(\mathcal{U}(\mathbf{e}_i), \mathcal{U}(\mathbf{e}_j)) = u_{1i}u_{1j} + u_{2i}u_{2j} + \dots + u_{ni}u_{nj}.$$

The expression on the right-hand side is equal to the element  $c_{ij}$ , where the matrix  $(c_{ij})$  is equal to  $U^*U$ . This implies that the condition of orthogonality of the transformation  $\mathcal{U}$  can be written in the form

$$U^*U = E, \quad (7.18)$$

or equivalently,  $U^* = U^{-1}$ . This equality is equivalent to

$$UU^* = E, \quad (7.19)$$

and can be expressed as relationships among the elements of the matrix  $U$ :

$$u_{i1}u_{j1} + \dots + u_{in}u_{jn} = 0 \quad \text{for } i \neq j, \quad u_{i1}^2 + \dots + u_{in}^2 = 1. \quad (7.20)$$

The matrix  $U$  satisfying the relationship (7.18) or the equivalent relationship (7.19) is said to be *orthogonal*.

The concept of an orthonormal basis of a Euclidean space can be interpreted more graphically using the notion of flag (see the definition on p. 101). Namely, we associate with an orthonormal basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  the flag

$$(\mathbf{0}) \subset L_1 \subset L_2 \subset \dots \subset L_n = L, \quad (7.21)$$

in which the subspace  $L_i$  is equal to  $\langle \mathbf{e}_1, \dots, \mathbf{e}_i \rangle$ , and the pair  $(L_{i-1}, L_i)$  is directed in the sense that  $L_i^+$  is the half-space of  $L_i$  containing the vector  $\mathbf{e}_i$ . In the case of a Euclidean space, the essential fact is that we obtain a *bijection* between orthonormal bases and flags.

For the proof of this, we have only to verify that the orthonormal basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  is uniquely determined by its associated flag. Let this basis be associated with the flag (7.21). If we have already constructed an orthonormal system of vectors  $\mathbf{e}_1, \dots, \mathbf{e}_{i-1}$  such that  $L_{i-1} = \langle \mathbf{e}_1, \dots, \mathbf{e}_{i-1} \rangle$ , then we should consider the orthogonal complement  $L_{i-1}^\perp$  of the subspace  $L_{i-1}$  in  $L_i$ . Then  $\dim L_{i-1}^\perp = 1$  and  $L_{i-1}^\perp = \langle \mathbf{e}_i \rangle$ , where the vector  $\mathbf{e}_i$  is uniquely defined up to the factor  $\pm 1$ . This factor can be selected unambiguously based on the condition  $\mathbf{e}_i \in L_i^+$ .

An observation made earlier can now be interpreted as follows: For any two flags  $\Phi_1$  and  $\Phi_2$  of a Euclidean space  $L$ , there exists a unique orthogonal transformation that maps  $\Phi_1$  to  $\Phi_2$ .

Our next goal will be the construction of an orthonormal basis in which a given orthogonal transformation  $\mathcal{U}$  has the simplest matrix possible. By Theorem 4.22, the transformation  $\mathcal{U}$  has a one- or two-dimensional invariant subspace  $L'$ . It is clear that the restriction of  $\mathcal{U}$  to the subspace  $L'$  is again an orthogonal transformation.

Let us determine first the sort of transformation that this can be, that is, what sorts of orthogonal transformations of one- and two-dimensional spaces exist.

If  $\dim L' = 1$ , then  $L' = \langle \mathbf{e} \rangle$  for some nonnull vector  $\mathbf{e}$ . Then  $\mathcal{U}(\mathbf{e}) = \alpha \mathbf{e}$ , where  $\alpha$  is some scalar. From the orthogonality of the transformation  $\mathcal{U}$ , we obtain that

$$(\mathbf{e}, \mathbf{e}) = (\alpha \mathbf{e}, \alpha \mathbf{e}) = \alpha^2 (\mathbf{e}, \mathbf{e}),$$

from which it follows that  $\alpha^2 = 1$ , and this implies that  $\alpha = \pm 1$ . Consequently, in a one-dimensional space  $L'$ , there exist two orthogonal transformations: the identity  $\mathcal{E}$ , for which  $\mathcal{E}(\mathbf{x}) = \mathbf{x}$  for all vectors  $\mathbf{x}$ , and the transformation  $\mathcal{U}$  such that  $\mathcal{U}(\mathbf{x}) = -\mathbf{x}$ . It is obvious that  $\mathcal{U} = -\mathcal{E}$ .

Now let  $\dim L' = 2$ , in which case  $L'$  is isomorphic to the plane with inner product (7.1). It is well known from analytic geometry that an orthogonal transformation of the plane is either a rotation through some angle  $\varphi$  about the origin or a reflection with respect to some line  $l$ . In the first case, the orthogonal transformation  $\mathcal{U}$  in an arbitrary orthonormal basis of the plane has matrix

$$\begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix}. \quad (7.22)$$

In the second case, the plane can be represented in the form of the direct sum  $L' = l \oplus l^\perp$ , where  $l$  and  $l^\perp$  are lines, and for a vector  $\mathbf{x}$  we have the decomposition  $\mathbf{x} = \mathbf{y} + \mathbf{z}$ , where  $\mathbf{y} \in l$  and  $\mathbf{z} \in l^\perp$ , while the vector  $\mathcal{U}(\mathbf{x})$  is equal to  $\mathbf{y} - \mathbf{z}$ . If we choose an orthonormal basis  $\mathbf{e}_1, \mathbf{e}_2$  in such a way that the vector  $\mathbf{e}_1$  lies on the line  $l$ , then the transformation  $\mathcal{U}$  will have matrix

$$U = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (7.23)$$

But we shall not presuppose this fact from analytic geometry, and instead show that it derives from simple considerations in linear algebra. Let  $\mathcal{U}$  have, in some orthonormal basis  $\mathbf{e}_1, \mathbf{e}_2$ , the matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad (7.24)$$

that is, it maps the vector  $x\mathbf{e}_1 + y\mathbf{e}_2$  to  $(ax + by)\mathbf{e}_1 + (cx + dy)\mathbf{e}_2$ . The fact that  $\mathcal{U}$  preserves the length of a vector gives the relationship

$$(ax + by)^2 + (cx + dy)^2 = x^2 + y^2$$

for all  $x$  and  $y$ . Substituting in turn  $(1, 0)$ ,  $(0, 1)$ , and  $(1, 1)$  for  $(x, y)$ , we obtain

$$a^2 + c^2 = 1, \quad b^2 + d^2 = 1, \quad ab + cd = 0. \quad (7.25)$$

From the relationship (7.19), it follows that  $|UU^*| = 1$ , and since  $|U^*| = |U|$ , it follows that  $|U|^2 = 1$ , and this implies that  $|U| = \pm 1$ . We need to consider separately the cases of different signs.

If  $|U| = -1$ , then the characteristic polynomial  $|U - tE|$  of the matrix (7.24) is equal to  $t^2 - (a + d)t - 1$  and has positive discriminant. Therefore, the matrix (7.24) has two real eigenvalues  $\lambda_1$  and  $\lambda_2$  of opposite signs (since by Viète's theorem,  $\lambda_1\lambda_2 = -1$ ) and two associated eigenvectors  $\mathbf{e}_1$  and  $\mathbf{e}_2$ . Examining the restriction of  $\mathcal{U}$  to the one-dimensional invariant subspaces  $\langle \mathbf{e}_1 \rangle$  and  $\langle \mathbf{e}_2 \rangle$ , we arrive at the one-dimensional case considered above, from which, in particular, it follows that the values  $\lambda_1$  and  $\lambda_2$  are equal to  $\pm 1$ . Let us show that the vectors  $\mathbf{e}_1$  and  $\mathbf{e}_2$  are orthogonal. By the definition of eigenvectors, we have the equalities  $\mathcal{U}(\mathbf{e}_i) = \lambda_i \mathbf{e}_i$ , from which we have

$$(\mathcal{U}(\mathbf{e}_1), \mathcal{U}(\mathbf{e}_2)) = (\lambda_1 \mathbf{e}_1, \lambda_2 \mathbf{e}_2) = \lambda_1 \lambda_2 (\mathbf{e}_1, \mathbf{e}_2). \quad (7.26)$$

But since the transformation  $\mathcal{U}$  is orthogonal, it follows that  $(\mathcal{U}(\mathbf{e}_1), \mathcal{U}(\mathbf{e}_2)) = (\mathbf{e}_1, \mathbf{e}_2)$ , and from (7.26), we obtain the equality  $(\mathbf{e}_1, \mathbf{e}_2) = \lambda_1 \lambda_2 (\mathbf{e}_1, \mathbf{e}_2)$ . Since  $\lambda_1$  and  $\lambda_2$  have opposite signs, it follows that  $(\mathbf{e}_1, \mathbf{e}_2) = 0$ . Choosing eigenvectors  $\mathbf{e}_1$  and  $\mathbf{e}_2$  of unit length and such that  $\lambda_1 = 1$  and  $\lambda_2 = -1$ , we obtain the orthonormal basis  $\mathbf{e}_1, \mathbf{e}_2$  in which the transformation  $\mathcal{U}$  has matrix (7.23). We then have the decomposition  $L = l \oplus l^\perp$ , where  $l = \langle \mathbf{e}_1 \rangle$  and  $l^\perp = \langle \mathbf{e}_2 \rangle$ , and the transformation  $\mathcal{U}$  is a reflection in the line  $l$ .

But if  $|U| = 1$ , then by relationship (7.25) for  $a, b, c, d$ , it is easy to derive, keeping in mind that  $ad - bc = 1$ , that there exists an angle  $\varphi$  such that  $a = d = \cos \varphi$  and  $c = -b = \sin \varphi$ , that is, the matrix (7.24) has the form (7.22).

As a basis for examining the general case, we have the following theorem.

**Theorem 7.24** *If a subspace  $L'$  is invariant with respect to an orthogonal transformation  $\mathcal{U}$ , then its orthogonal complement  $(L')^\perp$  is also invariant with respect to  $\mathcal{U}$ .*

*Proof* We must show that for every vector  $\mathbf{y} \in (L')^\perp$ , we have  $\mathcal{U}(\mathbf{y}) \in (L')^\perp$ . If  $\mathbf{y} \in (L')^\perp$ , then  $(\mathbf{x}, \mathbf{y}) = 0$  for all  $\mathbf{x} \in L'$ . From the orthogonality of the transformation  $\mathcal{U}$ , we obtain that  $(\mathcal{U}(\mathbf{x}), \mathcal{U}(\mathbf{y})) = (\mathbf{x}, \mathbf{y}) = 0$ . Since  $\mathcal{U}$  is a bijective mapping from  $L$  to  $L$ , its restriction to the invariant subspace  $L'$  is a bijection from  $L'$  to  $L'$ . In other words, every vector  $\mathbf{x}' \in L'$  can be represented in the form  $\mathbf{x}' = \mathcal{U}(\mathbf{x})$ , where  $\mathbf{x}$  is some other vector in  $L'$ . Consequently,  $(\mathbf{x}', \mathcal{U}(\mathbf{y})) = 0$  for every vector  $\mathbf{x}' \in L'$ , and this implies that  $\mathcal{U}(\mathbf{y}) \in (L')^\perp$ .  $\square$

**Remark 7.25** In the proof of Theorem 7.24, we nowhere used the positive definiteness of the quadratic form  $(\mathbf{x}, \mathbf{x})$  associated with the inner product  $(\mathbf{x}, \mathbf{y})$ . Indeed, this theorem holds as well for an arbitrary nonsingular bilinear form  $(\mathbf{x}, \mathbf{y})$ . The condition of nonsingularity is required in order that the restriction of the transformation  $\mathcal{U}$  to an invariant subspace be a bijection, without which the theorem would not be true.

**Definition 7.26** Subspaces  $L_1$  and  $L_2$  of a Euclidean space are said to be mutually *orthogonal* if  $(\mathbf{x}, \mathbf{y}) = 0$  for all vectors  $\mathbf{x} \in L_1$  and  $\mathbf{y} \in L_2$ . In such a case, we write

$L_1 \perp L_2$ . The decomposition of a Euclidean space as a direct sum of orthogonal subspaces is called an *orthogonal decomposition*.

If  $\dim L > 2$ , then by Theorem 4.22, the transformation  $\mathcal{U}$  has a one- or two-dimensional invariant subspace. Thus using Theorem 7.24 as many times as necessary (depending on  $\dim L$ ), we obtain the orthogonal decomposition

$$L = L_1 \oplus L_2 \oplus \cdots \oplus L_k, \quad \text{where } L_i \perp L_j \text{ for all } i \neq j, \quad (7.27)$$

with all subspaces  $L_i$  invariant with respect to the transformation  $\mathcal{U}$  and of dimension 1 or 2.

Combining the orthonormal bases of the subspaces  $L_1, \dots, L_k$  and choosing a convenient ordering, we obtain the following result.

**Theorem 7.27** *For every orthogonal transformation there exists an orthonormal basis in which the matrix of the transformation has the block-diagonal form*

$$\begin{pmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & & & & 0 & \\ & & & -1 & & & & \\ & & & & \ddots & & & \\ & & & & & -1 & & \\ & & & & & & A_{\varphi_1} & \\ & & 0 & & & & & \ddots \\ & & & & & & & & A_{\varphi_r} \end{pmatrix}, \quad (7.28)$$

where

$$A_{\varphi_i} = \begin{pmatrix} \cos \varphi_i & -\sin \varphi_i \\ \sin \varphi_i & \cos \varphi_i \end{pmatrix}, \quad (7.29)$$

$\varphi_i \neq \pi k, k \in \mathbb{Z}$ .

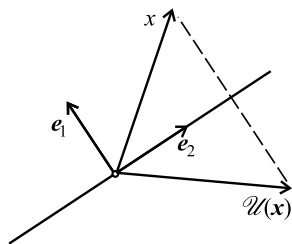
Let us note that the determinants of all the matrices (7.29) are equal to 1, and therefore, for a proper orthogonal transformation (see the definition on p. 135), the number of  $-1$ 's on the main diagonal in (7.28) is even, and for an improper orthogonal transformation, that number is odd.

Let us now look at what the theorems we have proved give us in the cases  $n = 1, 2, 3$  familiar from analytic geometry.

For  $n = 1$ , there exist, as we have already seen, altogether two orthogonal transformations, namely  $\mathcal{E}$  and  $-\mathcal{E}$ , the first of which is proper, and the second, improper.

For  $n = 2$ , a proper orthogonal transformation is a rotation of the plane through some angle  $\varphi$ . In an arbitrary orthonormal basis, its matrix has the form  $A_\varphi$  from (7.29), with no restriction on the angle  $\varphi$ . For the improper transformation appearing

**Fig. 7.3** Reflection of the plane with respect to a line



in (7.28), the number  $-1$  must be encountered an odd number of times, that is, once. This implies that in some orthonormal basis  $e_1, e_2$ , its matrix has the form

$$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This transformation is a reflection of the plane with respect to the line  $\langle e_2 \rangle$  (Fig. 7.3).

Let us now consider the case  $n = 3$ . Since the characteristic polynomial of the transformation  $\mathcal{U}$  has odd degree 3, it must have at least one real root. This implies that in the representation (7.28), the number  $+1$  or  $-1$  must appear on the main diagonal of the matrix.

Let us consider proper transformations first. In this case, for the matrix (7.28), we have only one possibility:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi & -\sin \varphi \\ 0 & \sin \varphi & \cos \varphi \end{pmatrix}.$$

If the matrix is written in the basis  $e_1, e_2, e_3$ , then the transformation  $\mathcal{U}$  does not change the points of the line  $l = \langle e_1 \rangle$  and represents a rotation through the angle  $\varphi$  in the plane  $\langle e_2, e_3 \rangle$ . In this case, we say that the transformation  $\mathcal{U}$  is a *rotation of the plane through the angle  $\varphi$  about the axis  $l$* . That every proper orthogonal transformation of a three-dimensional Euclidean space possesses a “rotational axis” is a result first proved by Euler. We shall discuss the mechanical significance of this assertion later, in connection with motions of affine spaces.

Finally, if an orthogonal transformation is improper, then in expression (7.28), we have only the possibility

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos \varphi & -\sin \varphi \\ 0 & \sin \varphi & \cos \varphi \end{pmatrix}.$$

In this case, the orthogonal transformation  $\mathcal{U}$  reduces to a rotation about the  $l$ -axis with a simultaneous *reflection* with respect to the plane  $l^\perp$ .

### 7.3 Orientation of a Euclidean Space\*

In a Euclidean space, as in any real vector space, there are defined the notions of equal and opposite *orientations* of two bases and *orientation* of the space (see Sect. 4.4). But in Euclidean spaces, these notions possess certain specific features.

Let  $e_1, \dots, e_n$  and  $e'_1, \dots, e'_n$  be two *orthonormal* bases of a Euclidean space  $L$ . By general definition, they have *equal orientations* if the transformation from one basis to the other is proper. This implies that for a transformation  $\mathcal{U}$  such that

$$\mathcal{U}(e_1) = e'_1, \quad \dots, \quad \mathcal{U}(e_n) = e'_n,$$

the determinant of its matrix is positive. But in the case that both bases under consideration are orthonormal, the mapping  $\mathcal{U}$ , as we know, is orthogonal, and its matrix  $U$  satisfies the relationship  $|U| = \pm 1$ . This implies that  $\mathcal{U}$  is a proper transformation if and only if  $|U| = 1$ , and it is improper if and only if  $|U| = -1$ . We have the following analogue to Theorems 4.38–4.40 of Sect. 4.4.

**Theorem 7.28** *Two orthogonal transformations of a real Euclidean space can be continuously deformed into each other if and only if the signs of their determinants coincide.*

The definition of a continuous deformation repeats here the definition given in Sect. 4.4 for the set  $\mathfrak{A}$ , but now consisting only of *orthogonal* matrices (or transformations). Since the product of any two orthogonal transformations is again orthogonal, Lemma 4.37 (p. 159) is also valid in this case, and we shall make use of it.

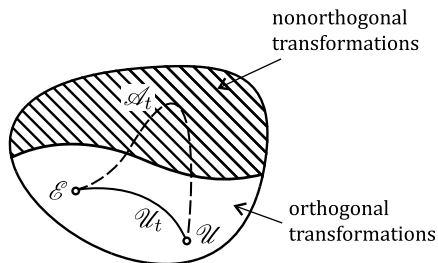
*Proof of Theorem 7.28* Let us show that an arbitrary proper orthogonal transformation  $\mathcal{U}$  can be continuously deformed into the identity. Since the condition of continuous deformability defines an equivalence relation on the set of orthogonal transformations, then by transitivity, the assertion of the theorem will follow for all proper transformations.

Thus we must prove that there exists a family of orthogonal transformations  $\mathcal{U}_t$  depending continuously on the parameter  $t \in [0, 1]$  for which  $\mathcal{U}_0 = \mathcal{E}$  and  $\mathcal{U}_1 = \mathcal{U}$ . The continuous dependence of  $\mathcal{U}_t$  implies that when it is represented in an arbitrary basis, all the elements of the matrices of the transformations  $\mathcal{U}_t$  are continuous functions of  $t$ . We note that this is not at all obvious corollary to Theorem 4.38. Indeed, it did not guarantee us that all the intermediate transformations  $\mathcal{U}_t$  for  $0 < t < 1$  are orthogonal. A possible “bad” deformation  $\mathcal{A}_t$  taking us out of the domain of orthogonal transformations is depicted as the dotted line in Fig. 7.4.

We shall use Theorem 7.27 and examine the orthonormal basis in which the matrix of the transformation  $\mathcal{U}$  has the form (7.28). The transformation  $\mathcal{U}$  is proper if and only if the number of instances of  $-1$  on the main diagonal of (7.28) is odd. We observe that the second-order matrix

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

**Fig. 7.4** Deformation taking us outside the domain of orthogonal transformations



can also be written in the form (7.29) for  $\varphi_i = \pi$ . Thus a proper orthogonal transformation can be written in a suitable orthonormal basis in block-diagonal form

$$\begin{pmatrix} E & & & \\ & A_{\varphi_1} & & \\ & & \ddots & \\ & & & A_{\varphi_k} \end{pmatrix}, \quad (7.30)$$

where the arguments  $\varphi_i$  can now be taken to be any values. Formula (7.30) in fact gives a continuous deformation of the transformation  $\mathcal{U}$  into  $\mathcal{E}$ . To maintain agreement with our notation, let us examine the transformations  $\mathcal{U}_t$  having in this same basis the matrix

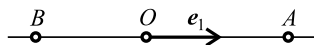
$$\begin{pmatrix} E & & & \\ & A_{t\varphi_1} & & \\ & & \ddots & \\ & & & A_{t\varphi_k} \end{pmatrix}. \quad (7.31)$$

Then it is clear first of all that the transformation  $\mathcal{U}_t$  is orthogonal for every  $t$ , and secondly, that  $\mathcal{U}_0 = \mathcal{E}$  and  $\mathcal{U}_1 = \mathcal{U}$ . This gives us a proof of the theorem in the case of a proper transformation.

Let us now consider improper orthogonal transformations and show that any such transformation  $\mathcal{V}$  can be continuously deformed into a reflection with respect to a hyperplane, that is, into a transformation  $\mathcal{F}$  having in some orthonormal basis the matrix

$$F = \begin{pmatrix} -1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix}. \quad (7.32)$$

Let us choose an arbitrary orthonormal basis of the vector space and suppose that in this basis, the improper orthogonal transformation  $\mathcal{V}$  has matrix  $V$ . Then it is obvious that the transformation  $\mathcal{U}$  with matrix  $U = VF$  in this same basis is a proper orthogonal transformation. Taking into account the obvious relationship  $F^{-1} = F$ , we have  $V = UF$ , that is,  $\mathcal{V} = \mathcal{U}\mathcal{F}$ . We shall use the family  $\mathcal{U}_t$  effecting a continuous deformation of the proper transformation  $\mathcal{U}$  into  $\mathcal{E}$ . From the preceding

**Fig. 7.5** Oriented length

equality, with the help of Lemma 4.37, we obtain the continuous family  $\mathcal{V}_t = \mathcal{U}_t \mathcal{F}$ , where  $\mathcal{V}_0 = \mathcal{E} \mathcal{F} = \mathcal{F}$  and  $\mathcal{V}_1 = \mathcal{U} \mathcal{F} = \mathcal{V}$ . Thus the family  $\mathcal{V}_t = \mathcal{U}_t \mathcal{F}$  effects the deformation of the improper transformation  $\mathcal{V}$  into  $\mathcal{F}$ .  $\square$

In analogy to what we did in Sect. 4.4, Theorem 7.28 gives us the following topological result: the set of orthogonal transformations consists of two path-connected components: the proper and improper orthogonal transformations.

Exactly as in Sect. 4.4, from what we have proved, it also follows that two equally oriented orthogonal bases can be continuously deformed into each other. That is, if  $\mathbf{e}_1, \dots, \mathbf{e}_n$  and  $\mathbf{e}'_1, \dots, \mathbf{e}'_n$  are orthogonal bases with the same orientation, then there exists a family of orthonormal bases  $\mathbf{e}_1(t), \dots, \mathbf{e}_n(t)$  depending continuously on the parameter  $t \in [0, 1]$  such that  $\mathbf{e}_i(0) = \mathbf{e}_i$  and  $\mathbf{e}_i(1) = \mathbf{e}'_i$ . In other words, the concept of orientation of a space is the same whether we define it in terms of an arbitrary basis or an orthonormal one. We shall further examine oriented Euclidean spaces, choosing an orientation arbitrarily. This choice makes it possible to speak of positively and negatively oriented orthonormal bases.

Now we can compare the concepts of oriented and unoriented volume. These two numbers differ by the factor  $\pm 1$  (unoriented volumes are nonnegative by definition). When the oriented volume of a parallelepiped  $\Pi(\mathbf{a}_1, \dots, \mathbf{a}_n)$  in a space  $L$  of dimension  $n$  was introduced, we noted that its definition depends on the choice of some orthonormal basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ . Since we are assuming that the space  $L$  is oriented, we can include in the definition of oriented volume of a parallelepiped  $\Pi(\mathbf{a}_1, \dots, \mathbf{a}_n)$  the condition that the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  used in the definition of  $v(\mathbf{a}_1, \dots, \mathbf{a}_n)$  be positively oriented. Then the number  $v(\mathbf{a}_1, \dots, \mathbf{a}_n)$  does not depend on the choice of basis (that is, it remains unchanged if instead of  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , we take any other orthonormal positively oriented basis  $\mathbf{e}'_1, \dots, \mathbf{e}'_n$ ). This follows immediately from formula (7.13) for the transformation  $\mathcal{C} = \mathcal{U}$  and from the fact that the transformation  $\mathcal{U}$  taking one basis to the other is orthogonal and proper, that is,  $|\mathcal{U}| = 1$ .

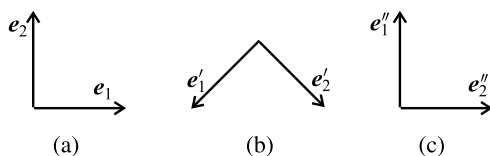
We can now say that the oriented volume  $v(\mathbf{a}_1, \dots, \mathbf{a}_n)$  is positive (and consequently equal to the unoriented volume) if the bases  $\mathbf{e}_1, \dots, \mathbf{e}_n$  and  $\mathbf{a}_1, \dots, \mathbf{a}_n$  are equally oriented, and is negative (that is, it differs from the unoriented volume by a sign) if these bases have opposite orientations. For example, on the line (Fig. 7.5), the length of the segment  $OA$  is equal to 2, while the length of the segment  $OB$  is equal to  $-2$ .

Thus, we may say that for the parallelepiped  $\Pi(\mathbf{a}_1, \dots, \mathbf{a}_n)$ , its oriented volume is its “volume with orientation.”

If we choose a coordinate origin on the real line, then a basis of it consists of a single vector, and vectors  $\mathbf{e}_1$  and  $\alpha \mathbf{e}_1$  are equally oriented if they lie to one side of the origin, that is,  $\alpha > 0$ . The choice of orientation on the line, one might say, corresponds to the choice of “right” and “left.”

In the real plane, the orientation given by the basis  $\mathbf{e}_1, \mathbf{e}_2$  is determined by the “direction of rotation” from  $\mathbf{e}_1$  to  $\mathbf{e}_2$ : clockwise or counterclockwise. Equally oriented bases  $\mathbf{e}_1, \mathbf{e}_2$  and  $\mathbf{e}'_1, \mathbf{e}'_2$  (Fig. 7.6(a) and (b)) can be continuously transformed



**Fig. 7.6** Oriented bases of the plane

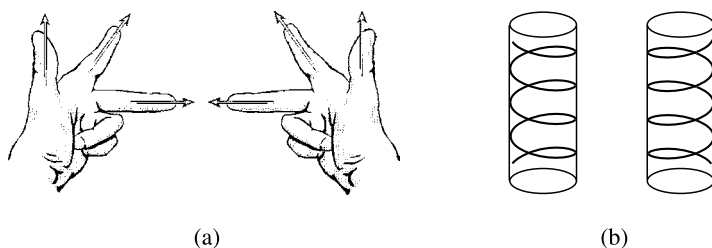
one into the other, while oppositely oriented bases cannot even if they form equal figures (Fig. 7.6(a) and (c)), since what is required for this is a reflection, that is, an improper transformation.

In real three-dimensional space, the orientation is defined by a basis of three orthonormal vectors. We again meet with two opposite orientations, which are represented by our right and left hands (see Fig. 7.7(a)). Another method of providing an orientation in three-dimensional space is defined by a helix (Fig. 7.7(b)). In this case, the orientation is defined by the direction in which the helix turns as it rises—clockwise or counterclockwise.<sup>2</sup>

## 7.4 Examples\*

**Example 7.29** By the term “figure” in a Euclidean space  $L$  we shall understand an arbitrary subset  $S \subset L$ . Two figures  $S$  and  $S'$  contained in a Euclidean space  $M$  of dimension  $n$  are said to be *congruent*, or *geometrically identical*, if there exists an orthogonal transformation  $\mathcal{U}$  of the space  $M$  taking  $S$  to  $S'$ . We shall be interested in the following question: When are figures  $S$  and  $S'$  congruent, that is, when do we have  $\mathcal{U}(S) = S'$ ?

Let us first deal with the case in which the figures  $S$  and  $S'$  consist of collections of  $m$  vectors:  $S = (\mathbf{a}_1, \dots, \mathbf{a}_m)$  and  $S' = (\mathbf{a}'_1, \dots, \mathbf{a}'_m)$  with  $m \leq n$ . For  $S$  and  $S'$  to be congruent is equivalent to the existence of an orthogonal transformation  $\mathcal{U}$  such that  $\mathcal{U}(\mathbf{a}_i) = \mathbf{a}'_i$  for all  $i = 1, \dots, m$ . For this, of course, it is necessary that the

**Fig. 7.7** Different orientations of three-dimensional space

<sup>2</sup>The molecules of amino acids likewise determine a certain orientation of space. In biology, the two possible orientations are designated by D (right = *dexter* in Latin) and L (left = *laevus*). For some unknown reason, they all determine the same orientation, namely the counterclockwise one.

following equality holds:

$$(\mathbf{a}_i, \mathbf{a}_j) = (\mathbf{a}'_i, \mathbf{a}'_j), \quad i, j = 1, \dots, m. \quad (7.33)$$

Let us assume that vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m$  are linearly independent, and we shall then prove that the condition (7.33) is sufficient. By Theorem 7.14, in this case we have  $G(\mathbf{a}_1, \dots, \mathbf{a}_m) > 0$ , and by assumption,  $G(\mathbf{a}'_1, \dots, \mathbf{a}'_m) = G(\mathbf{a}_1, \dots, \mathbf{a}_m)$ . From this same theorem, it follows that the vectors  $\mathbf{a}'_1, \dots, \mathbf{a}'_m$  will also be linearly independent.

Let us set

$$\mathbf{L} = \langle \mathbf{a}_1, \dots, \mathbf{a}_m \rangle, \quad \mathbf{L}' = \langle \mathbf{a}'_1, \dots, \mathbf{a}'_m \rangle, \quad (7.34)$$

and consider first the case  $m = n$ . Let  $\mathbf{M} = \langle \mathbf{a}_1, \dots, \mathbf{a}_m \rangle$ . We shall consider the transformation  $\mathcal{U} : \mathbf{M} \rightarrow \mathbf{M}$  given by the conditions  $\mathcal{U}(\mathbf{a}_i) = \mathbf{a}'_i$  for all  $i = 1, \dots, m$ . Obviously, such a transformation is uniquely determined, and by the relationship

$$\left( \mathcal{U} \left( \sum_{i=1}^m \alpha_i \mathbf{a}_i \right), \mathcal{U} \left( \sum_{j=1}^m \beta_j \mathbf{a}_j \right) \right) = \left( \sum_{i=1}^m \alpha_i \mathbf{a}'_i, \sum_{j=1}^m \beta_j \mathbf{a}'_j \right) = \sum_{i,j=1}^m \alpha_i \beta_j (\mathbf{a}'_i, \mathbf{a}'_j)$$

and equality (7.33), it is orthogonal.

Let  $m < n$ . Then we have the decomposition  $\mathbf{M} = \mathbf{L} \oplus \mathbf{L}^\perp = \mathbf{L}' \oplus (\mathbf{L}')^\perp$ , where the subspaces  $\mathbf{L}$  and  $\mathbf{L}'$  of the space  $\mathbf{M}$  are defined by formula (7.34). By what has gone before, there exists an isomorphism  $\mathcal{V} : \mathbf{L} \rightarrow \mathbf{L}'$  such that  $\mathcal{V}(\mathbf{a}_i) = \mathbf{a}'_i$  for all  $i = 1, \dots, m$ . The orthogonal complements  $\mathbf{L}^\perp$  and  $(\mathbf{L}')^\perp$  of these subspaces have dimension  $n - m$ , and consequently, are also isomorphic (Theorem 7.22). Let us choose an arbitrary isomorphism  $\mathcal{W} : \mathbf{L}^\perp \rightarrow (\mathbf{L}')^\perp$ . As a result of the decomposition  $\mathbf{M} = \mathbf{L} \oplus \mathbf{L}^\perp$ , an arbitrary vector  $\mathbf{x} \in \mathbf{M}$  can be uniquely represented in the form  $\mathbf{x} = \mathbf{y} + \mathbf{z}$ , where  $\mathbf{y} \in \mathbf{L}$  and  $\mathbf{z} \in \mathbf{L}^\perp$ . Let us define the linear transformation  $\mathcal{U} : \mathbf{M} \rightarrow \mathbf{M}$  by the formula  $\mathcal{U}(\mathbf{x}) = \mathcal{V}(\mathbf{y}) + \mathcal{W}(\mathbf{z})$ . By construction,  $\mathcal{U}(\mathbf{a}_i) = \mathbf{a}'_i$  for all  $i = 1, \dots, m$ , and a trivial verification shows that the transformation  $\mathcal{U}$  is orthogonal.

Let us now consider the case that  $S = l$  and  $S' = l'$  are lines, and consequently, consist of an infinite number of vectors. It suffices to set  $l = \langle \mathbf{e} \rangle$  and  $l' = \langle \mathbf{e}' \rangle$ , where  $|\mathbf{e}| = |\mathbf{e}'| = 1$ , and to use the fact that there exists an orthogonal transformation  $\mathcal{U}$  of the space  $\mathbf{M}$  taking  $\mathbf{e}$  to  $\mathbf{e}'$ . Thus any two lines are congruent.

The next case in order of increasing complexity is that in which figures  $S$  and  $S'$  each consist of two lines:  $S = l_1 \cup l_2$  and  $S' = l'_1 \cup l'_2$ . Let us set  $l_i = \langle \mathbf{e}_i \rangle$  and  $l'_i = \langle \mathbf{e}'_i \rangle$ , where  $|\mathbf{e}_i| = |\mathbf{e}'_i| = 1$  for  $i = 1$  and  $2$ . Now, however, vectors  $\mathbf{e}_1$  and  $\mathbf{e}_2$  are no longer defined uniquely, but can be replaced by  $-\mathbf{e}_1$  or  $-\mathbf{e}_2$ . In this case, their lengths do not change, but the inner product  $(\mathbf{e}_1, \mathbf{e}_2)$  can change their sign, that is, what remains unchanged is only their absolute value  $|(\mathbf{e}_1, \mathbf{e}_2)|$ . Based on previous considerations, we may say that figures  $S$  and  $S'$  are congruent if and only if  $|(\mathbf{e}_1, \mathbf{e}_2)| = |(\mathbf{e}'_1, \mathbf{e}'_2)|$ . If  $\varphi$  is the angle between the vectors  $\mathbf{e}_1$  and  $\mathbf{e}_2$ , then we see that the lines  $l_1$  and  $l_2$  determine  $|\cos \varphi|$ , or equivalently the angle  $\varphi$ , for which  $0 \leq \varphi \leq \frac{\pi}{2}$ . In textbooks on geometry, one often reads about two angles between straight lines, the “acute” and “obtuse” angles, but we shall choose only the one that

is acute or a right angle. This angle  $\varphi$  is called the *angle between the lines*  $l_1$  and  $l_2$ . The previous exposition shows that two pairs of lines  $l_1, l_2$  and  $l'_1, l'_2$  are congruent if and only if the angles between them thus defined coincide.

The case in which a figure  $S$  consists of a line  $l$  and a plane  $L$  ( $\dim l = 1$ ,  $\dim L = 2$ ) is also related, strictly speaking, to elementary geometry, since  $\dim(l + L) \leq 3$ , and the figure  $S = l \cup L$  can be embedded in three-dimensional space. But we shall consider it from a more abstract point of view, using the language of Euclidean spaces. Let  $l = \langle e \rangle$  and let  $f$  be the orthogonal projection of  $e$  onto  $L$ . The angle  $\varphi$  between the lines  $l$  and  $l' = \langle f \rangle$  is called the *angle between*  $l$  and  $L$  (as already mentioned above, it is acute or right). The cosine of this angle can be calculated according to the following formula:

$$\cos \varphi = \frac{|(e, f)|}{|e| \cdot |f|}. \quad (7.35)$$

Let us show that if the angle between the line  $l$  and the plane  $L$  is equal to the angle between the line  $l'$  and the plane  $L'$ , then the figures  $S = l \cup L$  and  $S' = l' \cup L'$  are congruent. First of all, it is obvious that there exists an orthogonal transformation taking  $L$  to  $L'$ , so that we may consider that  $L = L'$ . Let  $l = \langle e \rangle$ ,  $|e| = 1$  and  $l' = \langle e' \rangle$ ,  $|e'| = 1$ , and let us denote by  $f$  and  $f'$  the orthogonal projections  $e$  and  $e'$  onto  $L$ . By assumption,

$$\frac{|(e, f)|}{|e| \cdot |f|} = \frac{|(e', f')|}{|e'| \cdot |f'|}. \quad (7.36)$$

Since  $e$  and  $e'$  can be represented in the form  $e = f + x$  and  $e' = f' + y$ , where  $x, y \in L^\perp$ , it follows that  $|(e, f)| = |f|^2$ ,  $|(e', f')| = |f'|^2$ . Moreover,  $|e| = |e'| = 1$ , and the relationship (7.36) shows that  $|f| = |f'|$ .

Since  $e = x + f$ , we have  $|e|^2 = |x|^2 + 2(x, f) + |f|^2$ , from which, if we take into account the equalities  $|e|^2 = 1$  and  $(x, f) = 0$ , we obtain  $|x|^2 = 1 - |f|^2$  and analogously,  $|y|^2 = 1 - |f'|^2$ . From this follows the equality  $|x| = |y|$ . Let us define the orthogonal transformation  $\mathcal{U}$  of the space  $M = L \oplus L^\perp$  whose restriction to the plane  $L$  carries the vector  $f$  to  $f'$  (this is possible because  $|f| = |f'|$ ), while the restriction to its orthogonal complement  $L^\perp$  takes the vector  $x$  to  $y$  (which is possible on account of the equality  $|x| = |y|$ ). Clearly,  $\mathcal{U}$  takes  $e$  to  $e'$  and hence  $l$  to  $l'$ , and by construction, the plane  $L$  in both figures is one and the same, and the transformation  $\mathcal{U}$  takes it into itself.

We encounter a new and more interesting situation when we consider the case in which a figure  $S$  consists of a pair of planes  $L_1$  and  $L_2$  ( $\dim L_1 = \dim L_2 = 2$ ). If  $L_1 \cap L_2 \neq \{0\}$ , then  $\dim(L_1 + L_2) \leq 3$ , and we are dealing with a question from elementary geometry (which, however, can be considered simply in the language of Euclidean spaces). Therefore, we shall assume that  $L_1 \cap L_2 = \{0\}$  and similarly, that  $L'_1 \cap L'_2 = \{0\}$ . When are figures  $S = L_1 \cup L_2$  and  $S' = L'_1 \cup L'_2$  congruent? It turns out that for this to occur, it is necessary that there be agreement of not one (as in the examples considered above) but two parameters, which can be interpreted as *two angles* between the planes  $L_1$  and  $L_2$ .

We shall consider all possible straight lines lying in the plane  $L_1$  and the angles that they form with the plane  $L_2$ . To this end, we recall the geometric interpretation of the angle between a line  $l$  and a plane  $L$ . If  $l = \langle \mathbf{e} \rangle$ , where  $|\mathbf{e}| = 1$ , then the angle  $\varphi$  between  $l$  and  $L$  is determined by formula (7.35) with the condition  $0 \leq \varphi \leq \frac{\pi}{2}$ , where  $\mathbf{f}$  is the orthogonal projection of the vector  $\mathbf{e}$  onto  $L$ . From this, it follows that  $\mathbf{e} = \mathbf{f} + \mathbf{x}$ , where  $\mathbf{x} \in L^\perp$ , and this implies that  $(\mathbf{e}, \mathbf{f}) = (\mathbf{f}, \mathbf{f}) + (\mathbf{x}, \mathbf{f}) = |\mathbf{f}|^2$ , whence the relationship (7.35) gives  $|\cos \varphi| = |\mathbf{f}|$ . In other words, to consider all the angles between lines lying in the plane  $L_1$  and the plane  $L_2$ , we must consider the circle in  $L_1$  consisting of all vectors of length 1 and the lengths of the orthogonal projections of these vectors onto the plane  $L_2$ . In order to write down these angles in a formula, we shall consider the orthogonal projection  $M \rightarrow L_2$  of the space  $M$  onto the plane  $L_2$ . Let us denote by  $\mathcal{P}$  the restriction of this linear transformation onto the plane  $L_1$ . Then the angles of interest to us are given by the formula  $|\cos \varphi| = |\mathcal{P}(\mathbf{e})|$ , where  $\mathbf{e}$  are all possible vectors in the plane  $L_1$  of unit length. We restrict our attention to the case in which the linear transformation  $\mathcal{P}$  is an isomorphism. The case in which this does not occur, that is, when the kernel of the transformation  $\mathcal{P}$  is not equal to  $\{0\}$  and the image is not equal to  $L_2$ , is dealt with similarly.

Since  $\mathcal{P}$  is an isomorphism, there is an inverse transformation  $\mathcal{P}^{-1}: L_2 \rightarrow L_1$ . Let us choose in the planes  $L_1$  and  $L_2$  orthonormal bases  $\mathbf{e}_1, \mathbf{e}_2$  and  $\mathbf{g}_1, \mathbf{g}_2$ . Let the vector  $\mathbf{e} \in L_1$  have unit length. We set  $\mathbf{f} = \mathcal{P}(\mathbf{e})$ , and assuming that  $\mathbf{f} = x_1 \mathbf{g}_1 + x_2 \mathbf{g}_2$ , we shall obtain equations for the coordinates  $x_1$  and  $x_2$ . Let us set

$$\mathcal{P}^{-1}(\mathbf{g}_1) = \alpha \mathbf{e}_1 + \beta \mathbf{e}_2, \quad \mathcal{P}^{-1}(\mathbf{g}_2) = \gamma \mathbf{e}_1 + \delta \mathbf{e}_2.$$

Since  $\mathbf{f} = \mathcal{P}(\mathbf{e})$ , it follows that

$$\mathbf{e} = \mathcal{P}^{-1}(\mathbf{f}) = x_1 \mathcal{P}^{-1}(\mathbf{g}_1) + x_2 \mathcal{P}^{-1}(\mathbf{g}_2) = (\alpha x_1 + \gamma x_2) \mathbf{e}_1 + (\beta x_1 + \delta x_2) \mathbf{e}_2,$$

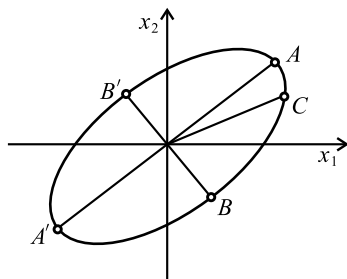
and the condition  $|\mathcal{P}^{-1}(\mathbf{f})| = 1$ , which we shall write in the form  $|\mathcal{P}^{-1}(\mathbf{f})|^2 = 1$ , reduces to the equality  $(\alpha x_1 + \gamma x_2)^2 + (\beta x_1 + \delta x_2)^2 = 1$ , that is,

$$(\alpha^2 + \beta^2)x_1^2 + 2(\alpha\gamma + \beta\delta)x_1x_2 + (\gamma^2 + \delta^2)x_2^2 = 1. \quad (7.37)$$

Equation (7.37) with variables  $x_1, x_2$  defines a second-degree curve in the rectangular coordinate system determined by the vectors  $\mathbf{g}_1$  and  $\mathbf{g}_2$ . This curve is bounded, since  $|\mathbf{f}| \leq |\mathbf{e}|$  ( $\mathbf{f}$  is the orthogonal projection of the vector  $\mathbf{e}$ ), and this implies that  $(\mathbf{f}^2) \leq 1$ , that is,  $x_1^2 + x_2^2 \leq 1$ . As one learns in a course on analytic geometry, such a curve is an ellipse. In our case, it has its center of symmetry at the origin  $O$ , that is, it is unchanged by a change of variables  $x_1 \rightarrow -x_1, x_2 \rightarrow -x_2$  (see Fig. 7.8).

It is known from analytic geometry that an ellipse has two distinguished points  $A$  and  $A'$ , symmetric with respect to the origin, such that the length  $|OA| = |OA'|$  is greater than  $|OC|$  for all other points  $C$  of the ellipse. The segment  $|OA| = |OA'|$  is called the *semimajor axis* of the ellipse. Similarly, there exist points  $B$  and  $B'$  symmetric with respect to the origin such that the segment  $|OB| = |OB'|$  is shorter than every other segment  $|OC|$ . The segment  $|OB| = |OB'|$  is called the *semiminor axis* of the ellipse.

**Fig. 7.8** *Ellipse described by equation (7.37)*



Let us recall that the length of an arbitrary line segment  $|OC|$ , where  $C$  is any point on the ellipse, gives us the value  $\cos \varphi$ , where  $\varphi$  is the angle between a certain line contained in  $L_1$  and the plane  $L_2$ . From this it follows that  $\cos \varphi$  attains its maximum for one value of  $\varphi$ , while for some other value of  $\varphi$  it attains its minimum. Let us denote these angles by  $\varphi_1$  and  $\varphi_2$  respectively. By definition,  $0 \leq \varphi_1 \leq \varphi_2 \leq \frac{\pi}{2}$ . It is these *two* angles that are called the *angles between the planes  $L_1$  and  $L_2$* .

The case that we have omitted, in which the transformation  $\mathcal{P}$  has a nonnull kernel, reduces to the case in which the ellipse depicted in Fig. 7.8 shrinks to a line segment.

It now remains for us to check that if both angles between the planes  $(L_1, L_2)$  are equal to the corresponding angles between the planes  $(L'_1, L'_2)$ , then the figures  $S = L_1 \cup L_2$  and  $S' = L'_1 \cup L'_2$  will be congruent, that is, there exists an orthogonal transformation  $\mathcal{U}$  taking the plane  $L_i$  into  $L'_i$ ,  $i = 1, 2$ .

Let  $\varphi_1$  and  $\varphi_2$  be the angles between  $L_1$  and  $L_2$ , equal, by hypothesis, to the angles between  $L'_1$  and  $L'_2$ . Reasoning as previously (in the case of the angle between a line and a plane), we can find an orthogonal transformation that takes  $L_2$  to  $L'_2$ . This implies that we may assume that  $L_2 = L'_2$ . Let us denote this plane by  $L$ . Here, of course, the angles  $\varphi_1$  and  $\varphi_2$  remain unchanged. Let  $\cos \varphi_1 \leq \cos \varphi_2$  for the pair of planes  $L_1$  and  $L$ . This implies that  $\cos \varphi_1$  and  $\cos \varphi_2$  are the lengths of the semiminor and semimajor axes of the ellipse that we considered above. This is also the case for the pair of planes  $L'_1$  and  $L$ . By construction, this means that  $\cos \varphi_1 = |f_1| = |f'_1|$  and  $\cos \varphi_2 = |f_2| = |f'_2|$ , where the vectors  $f_i \in L$  are orthogonal projections of the vectors  $e_i \in L_1$  of length 1. Reasoning similarly, we obtain the vectors  $f'_i \in L$  and  $e'_i \in L'_1$ ,  $i = 1, 2$ .

Since  $|f_1| = |f'_1|$ ,  $|f_2| = |f'_2|$ , and since by well-known properties of the ellipse, its semimajor and semiminor axes are orthogonal, we can find an orthogonal transformation of the space  $M$  that takes  $f_1$  to  $f'_1$  and  $f_2$  to  $f'_2$ , and having done so, assume that  $f_1 = f'_1$  and  $f_2 = f'_2$ . But since an ellipse is defined by its semiaxes, it follows that the ellipses  $C_1$  and  $C'_1$  that are obtained in the plane  $L$  from the planes  $L_1$  and  $L'_1$  simply coincide. Let us consider the orthogonal projections of the space  $M$  to the plane  $L$ . Let us denote by  $\mathcal{P}$  its restriction to the plane  $L_1$ , and by  $\mathcal{P}'$  its restriction to the plane  $L'_1$ .

We shall assume, as we did previously, that the transformations  $\mathcal{P} : L_1 \rightarrow L$  and  $\mathcal{P}' : L'_1 \rightarrow L$  are isomorphisms of the corresponding linear spaces, but it is not at all necessary that they be isomorphisms of Euclidean spaces. Let us represent this with

arrows in a commutative diagram

$$\begin{array}{ccc}
 L_1 & & \\
 \downarrow \mathcal{V} & \searrow \mathcal{P} & \\
 & L & \\
 & \nearrow \mathcal{P}' & \\
 L'_1 & &
 \end{array} \tag{7.38}$$

and let us show that the transformations  $\mathcal{P}$  and  $\mathcal{P}'$  differ from each other by an isomorphism of Euclidean spaces  $L_1$  and  $L'_1$ . In other words, we claim that the transformation  $\mathcal{V} = (\mathcal{P}')^{-1}\mathcal{P}$  is an isomorphism of the Euclidean spaces  $L_1$  and  $L'_1$ .

As the product of isomorphisms of linear spaces, the transformation  $\mathcal{V}$  is also an isomorphism, that is, a bijective linear transformation. It remains for us to verify that  $\mathcal{V}$  preserves the inner product. As noted above, to do this, it suffices to verify that  $\mathcal{V}$  preserves the lengths of vectors. Let  $\mathbf{x}$  be a vector in  $L$ . If  $\mathbf{x} = \mathbf{0}$ , then the vector  $\mathcal{V}(\mathbf{x})$  is equal to  $\mathbf{0}$  by the linearity of  $\mathcal{V}$ , and the assertion is obvious. If  $\mathbf{x} \neq \mathbf{0}$ , then we set  $\mathbf{e} = \alpha^{-1}\mathbf{x}$ , where  $\alpha = |\mathbf{x}|$ , and then  $|\mathbf{e}| = 1$ . The vector  $\mathcal{P}(\mathbf{e})$  is contained in the ellipse  $C$  in the plane  $L$ . Since  $C = C'$ , it follows that  $\mathcal{P}(\mathbf{e}) = \mathcal{P}'(\mathbf{e}')$ , where  $\mathbf{e}'$  is some vector in the plane  $L'_1$  and  $|\mathbf{e}'| = 1$ . From this we obtain the equality  $(\mathcal{P}')^{-1}\mathcal{P}(\mathbf{e}) = \mathbf{e}'$ , that is,  $\mathcal{V}(\mathbf{e}) = \mathbf{e}'$  and  $|\mathbf{e}'| = 1$ , which implies that  $|\mathcal{V}(\mathbf{x})| = \alpha = |\mathbf{x}|$ , which is what we had to prove.

We shall now consider a basis of the plane  $L$  consisting of vectors  $\mathbf{f}_1$  and  $\mathbf{f}_2$  lying on the semimajor and semiminor axes of the ellipse  $C = C'$ , and augment it with vectors  $\mathbf{e}_1, \mathbf{e}_2$ , where  $\mathcal{P}(\mathbf{e}_i) = \mathbf{f}_i$ . We thereby obtain four vectors  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{f}_1, \mathbf{f}_2$  in the space  $L_1 + L$  (it is easily verified that they are linearly independent). Similarly, in the space  $L'_1 + L$ , we shall construct four vectors  $\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{f}_1, \mathbf{f}_2$ . We shall show that there exists an orthogonal transformation of the space  $M$  taking the first set of four vectors into the second. To do so, it suffices to prove that the inner products of the associated vectors (in the order in which we have written them) coincide. Here what is least trivial is the relationship  $(\mathbf{e}'_1, \mathbf{e}'_2) = (\mathbf{e}_1, \mathbf{e}_2)$ , but it follows from the fact that  $\mathbf{e}'_i = \mathcal{V}(\mathbf{e}_i)$ , where  $\mathcal{V}$  is an isomorphism of the Euclidean spaces  $L_1$  and  $L'_1$ . The relationship  $(\mathbf{e}'_1, \mathbf{f}_1) = (\mathbf{e}_1, \mathbf{f}_1)$  is a consequence of the fact that  $\mathbf{f}_1$  is an orthogonal projection,  $(\mathbf{e}_1, \mathbf{f}_1) = |\mathbf{f}_1|^2$ , and similarly,  $(\mathbf{e}'_1, \mathbf{f}_1) = |\mathbf{f}_1|^2$ . The remaining relationships are even more obvious.

Thus the figures  $S = L_1 \cup L_2$  and  $S' = L'_1 \cup L'_2$  are congruent if and only if *both* angles between the planes  $L_1, L_2$  and  $L'_1, L'_2$  coincide. With the help of theorems to be proved in Sect. 7.5, it will be easy for the reader to investigate the case of a pair of subspaces  $L_1, L_2 \subset M$  of arbitrary dimension. In this case, the answer to the question whether two pairs of subspaces  $S = L_1 \cup L_2$  and  $S' = L'_1 \cup L'_2$  are congruent is determined by the agreement of two finite sets of numbers that can be interpreted as “angles” between the subspaces  $L_1, L_2$  and  $L'_1, L'_2$ .

*Example 7.30* When the senior of the two authors of this textbook gave the course on which it is based (this was probably in 1952 or 1953) at Moscow State University, he told his students about a question that had arisen in the work of A.N. Kolmogorov, A.A. Petrov, and N.V. Smirnov, the answer to which in one particular case had been obtained by A.I. Maltsev. This question was presented by the professor as an example of an unsolved problem that had been worked on by noted mathematicians yet could be formulated entirely in the language of linear algebra. At the next lecture, that is, a week later, one of the students in the class came up to him and said that he had found a solution to the problem.<sup>3</sup>

The question posed by A.N. Kolmogorov et al. was this: In a Euclidean space  $L$  of dimension  $n$ , we are given  $n$  nonnull mutually orthogonal vectors  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , that is,  $(\mathbf{x}_i, \mathbf{x}_j) = 0$  for all  $i \neq j$ ,  $i, j = 1, \dots, n$ . For what values  $m < n$  does there exist an  $m$ -dimensional subspace  $M \subset L$  such that the orthogonal projections of the vectors  $\mathbf{x}_1, \dots, \mathbf{x}_n$  to it all have the same length? A.I. Maltsev showed that if all the vectors  $\mathbf{x}_1, \dots, \mathbf{x}_n$  have the same length, then there exists such a subspace  $M$  of each dimension  $m < n$ .

The general case is approached as follows. Let us set  $|\mathbf{x}_i| = \alpha_i$  and assume that there exists an  $m$ -dimensional subspace  $M$  such that the orthogonal projections of all vectors  $\mathbf{x}_i$  to it have the same length  $\alpha$ . Let us denote by  $\mathcal{P}$  the orthogonal mapping to the subspace  $M$ , so that  $|\mathcal{P}(\mathbf{x}_i)| = \alpha$ . Let us set  $\mathbf{f}_i = \alpha_i^{-1} \mathbf{x}_i$ . Then the vectors  $\mathbf{f}_1, \dots, \mathbf{f}_n$  form an orthonormal basis of the space  $L$ . Conversely, let us select in  $L$  an orthonormal basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  such that the vectors  $\mathbf{e}_1, \dots, \mathbf{e}_m$  form a basis in  $M$ , that is, for the decomposition

$$L = M \oplus M^\perp, \quad (7.39)$$

we join the orthonormal basis  $\mathbf{e}_1, \dots, \mathbf{e}_m$  of the subspace  $M$  to the orthonormal basis  $\mathbf{e}_{m+1}, \dots, \mathbf{e}_n$  of the subspace  $M^\perp$ .

Let  $\mathbf{f}_i = \sum_{k=1}^n u_{ki} \mathbf{e}_k$ . Then we can interpret the matrix  $U = (u_{ki})$  as the matrix of the linear transformation  $\mathcal{U}$ , written in terms of the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , taking vectors  $\mathbf{e}_1, \dots, \mathbf{e}_n$  to vectors  $\mathbf{f}_1, \dots, \mathbf{f}_n$ . Since both sets of vectors  $\mathbf{e}_1, \dots, \mathbf{e}_n$  and  $\mathbf{f}_1, \dots, \mathbf{f}_n$  are orthonormal bases, it follows that  $\mathcal{U}$  is an orthogonal transformation, in particular, by formula (7.18), satisfying the relationship

$$UU^* = E. \quad (7.40)$$

From the decomposition (7.39) we see that every vector  $\mathbf{f}_i$  can be uniquely represented in the form of a sum  $\mathbf{f}_i = \mathbf{u}_i + \mathbf{v}_i$ , where  $\mathbf{u}_i \in M$  and  $\mathbf{v}_i \in M^\perp$ . By definition, the orthogonal projection of the vector  $\mathbf{f}_i$  onto the subspace  $M$  is equal to  $\mathcal{P}(\mathbf{f}_i) = \mathbf{u}_i$ . By construction of the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , it follows that

$$\mathcal{P}(\mathbf{f}_i) = \sum_{k=1}^m u_{ki} \mathbf{e}_k.$$

<sup>3</sup>It was published as L.B. Nisnevich, V.I. Bryzgalov, "On a problem of  $n$ -dimensional geometry," *Uspekhi Mat. Nauk* 8:4(56) (1953), 169–172.

By assumption, we have the equalities  $|\mathcal{P}(f_i)|^2 = |\mathcal{P}(\alpha_i^{-1}x_i)|^2 = \alpha^2\alpha_i^{-2}$ , which in coordinates assume the form

$$\sum_{k=1}^m u_{ki}^2 = \alpha^2\alpha_i^{-2}, \quad i = 1, \dots, n.$$

If we sum these relationships for all  $i = 1, \dots, n$  and change the order of summation in the double sum, then taking into account the relationship (7.40) for the orthogonal matrix  $U$ , we obtain the equality

$$\alpha^2 \sum_{i=1}^n \alpha_i^{-2} = \sum_{i=1}^n \sum_{k=1}^m u_{ki}^2 = \sum_{k=1}^m \sum_{i=1}^n u_{ki}^2 = m, \quad (7.41)$$

from which it follows that  $\alpha$  can be expressed in terms of  $\alpha_1, \dots, \alpha_n$ , and  $m$  by the formula

$$\alpha^2 = m \left( \sum_{i=1}^n \alpha_i^{-2} \right)^{-1}. \quad (7.42)$$

From this, in view of the equalities  $|\mathcal{P}(f_i)|^2 = |\mathcal{P}(\alpha_i^{-1}x_i)|^2 = \alpha^2\alpha_i^{-2}$ , we obtain the expressions

$$|\mathcal{P}(f_i)|^2 = m \left( \alpha_i^2 \sum_{i=1}^n \alpha_i^{-2} \right)^{-1}, \quad i = 1, \dots, n.$$

By Theorem 7.10, we have  $|\mathcal{P}(f_i)| \leq |f_i|$ , and since by construction,  $|f_i| = 1$ , we obtain the inequalities

$$m \left( \alpha_i^2 \sum_{i=1}^n \alpha_i^{-2} \right)^{-1} \leq 1, \quad i = 1, \dots, n,$$

from which it follows that

$$\alpha_i^2 \sum_{i=1}^n \alpha_i^{-2} \geq m, \quad i = 1, \dots, n. \quad (7.43)$$

Thus the inequalities (7.43) are *necessary* for the solvability of the problem. Let us show that they are also *sufficient*.

Let us consider first the case  $m = 1$ . We observe that in this situation, the inequalities (7.43) are automatically satisfied for an arbitrary collection of positive numbers  $\alpha_1, \dots, \alpha_n$ . Therefore, for an arbitrary system of mutually orthogonal vectors  $x_1, \dots, x_n$  in  $L$ , we must produce a line  $M \subset L$  such that the orthogonal projections of all these vectors onto it have the same length. For this, we shall take as such



a line  $M = \langle \mathbf{y} \rangle$  with the vectors

$$\mathbf{y} = \sum_{i=1}^n \frac{(\alpha_1 \cdots \alpha_n)^2}{\alpha_i^2} \mathbf{x}_i,$$

where as before,  $\alpha_i^2 = (\mathbf{x}_i, \mathbf{x}_i)$ . Since  $\frac{(\mathbf{x}_i, \mathbf{y})}{|\mathbf{y}|^2} \mathbf{y} \in M$  and  $(\mathbf{x}_i - \frac{(\mathbf{x}_i, \mathbf{y})}{|\mathbf{y}|^2} \mathbf{y}, \mathbf{y}) = 0$ , it follows that the orthogonal projection of the vector  $\mathbf{x}_i$  onto the line  $M$  is equal to

$$\mathcal{P}(\mathbf{x}_i) = \frac{(\mathbf{x}_i, \mathbf{y})}{|\mathbf{y}|^2} \mathbf{y}.$$

Clearly, the length of each such projection

$$|\mathcal{P}(\mathbf{x}_i)| = \frac{|(\mathbf{x}_i, \mathbf{y})|}{|\mathbf{y}|} = \frac{(\alpha_1 \cdots \alpha_n)^2}{|\mathbf{y}|}$$

does not depend on the index of the vector  $\mathbf{x}_i$ . Thus we have proved that for an arbitrary system of  $n$  nonnull mutually orthogonal vectors in an  $n$ -dimensional Euclidean space, there exists a line such that the orthogonal projections of all vectors onto it have the same length.

To facilitate understanding in what follows, we shall use the symbol  $P(m, n)$  to denote the following assertion: If the lengths  $\alpha_1, \dots, \alpha_n$  of a system of mutually orthogonal vectors  $\mathbf{x}_1, \dots, \mathbf{x}_n$  in an  $n$ -dimensional Euclidean space  $L$  satisfy condition (7.43), then there exists an  $m$ -dimensional subspace  $M \subset L$  such that the orthogonal projections  $\mathcal{P}(\mathbf{x}_1), \dots, \mathcal{P}(\mathbf{x}_n)$  of the vectors  $\mathbf{x}_1, \dots, \mathbf{x}_n$  onto it have the same length  $\alpha$ , expressed by the formula (7.42). Using this convention, we may say that we have proved the assertion  $P(1, n)$  for all  $n > 1$ .

Before passing to the case of arbitrary  $m$ , let us recast the problem in a more convenient form. Let  $\beta_1, \dots, \beta_n$  be arbitrary numbers satisfying the following condition:

$$\beta_1 + \cdots + \beta_n = m, \quad 0 < \beta_i \leq 1, i = 1, \dots, n. \quad (7.44)$$

Let us denote by  $P'(m, n)$  the following assertion: In the Euclidean space  $L$  there exist an orthonormal basis  $\mathbf{g}_1, \dots, \mathbf{g}_n$  and an  $m$ -dimensional subspace  $L' \subset L$  such that the orthogonal projections  $\mathcal{P}'(\mathbf{g}_i)$  of the basis vectors onto  $L'$  have length  $\sqrt{\beta_i}$ , that is,

$$|\mathcal{P}'(\mathbf{g}_i)|^2 = \beta_i, \quad i = 1, \dots, n.$$

**Lemma 7.31** *The assertions  $P(m, n)$  and  $P'(m, n)$  with a suitable choice of numbers  $\alpha_1, \dots, \alpha_n$  and  $\beta_1, \dots, \beta_n$  are equivalent.*

*Proof* Let us first prove that the assertion  $P'(m, n)$  follows from the assertion  $P(m, n)$ . Here we are given a collection of numbers  $\beta_1, \dots, \beta_n$  satisfying the condition (7.44), and it is known that the assertion  $P(m, n)$  holds for arbitrary positive

numbers  $\alpha_1, \dots, \alpha_n$  satisfying condition (7.43). For the numbers  $\beta_1, \dots, \beta_n$  and arbitrary orthonormal basis  $\mathbf{g}_1, \dots, \mathbf{g}_n$  we define vectors  $\mathbf{x}_i = \beta_i^{-1/2} \mathbf{g}_i$ ,  $i = 1, \dots, n$ . It is clear that these vectors are mutually orthogonal, and furthermore,  $|\mathbf{x}_i| = \beta_i^{-1/2}$ . Let us prove that the numbers  $\alpha_i = \beta_i^{-1/2}$  satisfy the inequalities (7.43). Indeed, if we take into account the condition (7.44), we have

$$\alpha_i^2 \sum_{i=1}^n \alpha_i^{-2} = \beta_i^{-1} \sum_{i=1}^n \beta_i = \beta_i^{-1} m \geq m.$$

The assertion  $P(m, n)$  says that in the space  $L$  there exists an  $m$ -dimensional subspace  $M$  such that the lengths of the orthogonal projections of the vectors  $\mathbf{x}_i$  onto it are equal to

$$|\mathcal{P}(\mathbf{x}_i)| = \alpha = \sqrt{m \left( \sum_{i=1}^n \alpha_i^{-2} \right)^{-1}} = \sqrt{m \left( \sum_{i=1}^n \beta_i \right)^{-1}} = 1.$$

But then the lengths of the orthogonal projections of the vectors  $\mathbf{g}_i$  onto the same subspace  $M$  are equal to  $|\mathcal{P}(\mathbf{g}_i)| = |\mathcal{P}(\sqrt{\beta_i} \mathbf{x}_i)| = \sqrt{\beta_i}$ .

Now let us prove that the assertion  $P'(m, n)$  yields  $P(m, n)$ . Here we are given a collection of nonnull mutually orthogonal vectors  $\mathbf{x}_1, \dots, \mathbf{x}_n$  of length  $|\mathbf{x}_i| = \alpha_i$ , and moreover, the numbers  $\alpha_i$  satisfy the inequalities (7.43). Let us set

$$\beta_i = \alpha_i^{-2} m \left( \sum_{i=1}^n \alpha_i^{-2} \right)^{-1}$$

and verify that  $\beta_i$  satisfies conditions (7.44). The equality  $\beta_1 + \dots + \beta_n = m$  clearly follows from the definition of the numbers  $\beta_i$ . From the inequalities (7.43) it follows that

$$\alpha_i^2 \geq \left( m \sum_{i=1}^n \alpha_i^{-2} \right)^{-1},$$

and this implies that

$$\beta_i = \alpha_i^{-2} m \left( \sum_{i=1}^n \alpha_i^{-2} \right)^{-1} \leq 1.$$

The assertion  $P'(m, n)$  says that there exist an orthonormal basis  $\mathbf{g}_1, \dots, \mathbf{g}_n$  of the space  $L$  and an  $m$ -dimensional subspace  $L' \subset L$  such that the lengths of the orthogonal projections of the vectors  $\mathbf{g}_i$  onto it are equal to  $|\mathcal{P}'(\mathbf{g}_i)| = \sqrt{\beta_i}$ . But then the orthogonal projections of the mutually orthogonal vectors  $\beta_i^{-1/2} \mathbf{g}_i$  onto the same subspace  $L'$  will have the same length, namely 1.

To prove the assertion  $P(m, n)$  for given vectors  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , it now suffices to consider the linear transformation  $\mathcal{U}$  of the space  $L$  mapping the vectors  $\mathbf{g}_i$  to

$\mathcal{U}(\mathbf{g}_i) = \mathbf{f}_i$ , where  $\mathbf{f}_i = \alpha_i^{-1} \mathbf{x}_i$ . Since the bases  $\mathbf{g}_1, \dots, \mathbf{g}_n$  and  $\mathbf{f}_1, \dots, \mathbf{f}_n$  are orthonormal, it follows that  $\mathcal{U}$  is an orthogonal transformation, and therefore, the orthogonal projections of the  $\mathbf{x}_i$  onto the  $m$ -dimensional subspace  $\mathbf{M} = \mathcal{U}(\mathbf{L}')$  have the same length. Moreover, by what we have proved above, this length is equal to the number  $\alpha$  determined by formula (7.42). This completes the proof of the lemma.  $\square$

Thanks to the lemma, we may prove the assertion  $P'(m, n)$  instead of the assertion  $P(m, n)$ . We shall do so by induction on  $m$  and  $n$ . We have already proved the base case of the induction ( $m = 1, n > 1$ ). The inductive step will be divided into three parts:

- (1) From assertion  $P'(m, n)$  for  $2m \leq n + 1$  we shall derive  $P'(m, n + 1)$ .
- (2) We shall prove that the assertion  $P'(m, n)$  implies  $P'(n, m - n)$ .
- (3) We shall prove that the assertion  $P'(m + 1, n)$  for all  $n > m + 1$  is a consequence of the assertion  $P'(m', n)$  for all  $m' \leq m$  and  $n > m'$ .

*Part 1:* From assertion  $P'(m, n)$  for  $2m \leq n + 1$ , we derive  $P'(m, n + 1)$ . We shall consider the collection of positive numbers  $\beta_1, \dots, \beta_n, \beta_{n+1}$  satisfying conditions (7.44) with  $n$  replaced by  $n + 1$ , with  $2m \leq (n + 1)$ . Without loss of generality, we may assume that  $\beta_1 \geq \beta_2 \geq \dots \geq \beta_{n+1}$ . Since  $\beta_1 + \dots + \beta_{n+1} = m$  and  $n + 1 \geq 2m$ , it follows that  $\beta_n + \beta_{n+1} \leq 1$ . Indeed, for example for odd  $n$ , the contrary assumption would give the inequality

$$\underbrace{\beta_1 + \beta_2 \geq \dots \geq \beta_n + \beta_{n+1}}_{(n+1)/2 \text{ sums}} > 1,$$

from which clearly follows  $\beta_1 + \dots + \beta_{n+1} > (n + 1)/2 \geq m$ , which contradicts the assumption that has been made.

Let us consider the  $(n + 1)$ -dimensional Euclidean space  $\mathbf{L}$  and decompose it as a direct sum  $\mathbf{L} = \langle \mathbf{e} \rangle \oplus \langle \mathbf{e} \rangle^\perp$ , where  $\mathbf{e} \in \mathbf{L}$  is an arbitrary vector of length 1. By the induction hypothesis, the assertion  $P'(m, n)$  holds for numbers  $\beta_1, \dots, \beta_{n-1}$  and  $\beta = \beta_n + \beta_{n+1}$  and the  $n$ -dimensional Euclidean space  $\langle \mathbf{e} \rangle^\perp$ . This implies that in the space  $\langle \mathbf{e} \rangle^\perp$ , there exist an orthonormal basis  $\mathbf{g}_1, \dots, \mathbf{g}_n$  and an  $m$ -dimensional subspace  $\mathbf{L}'$  such that the squares of the lengths of the orthogonal projections of the vectors  $\mathbf{g}_i$  onto  $\mathbf{L}'$  are equal to

$$|\mathcal{P}'(\mathbf{g}_i)|^2 = \beta_i, \quad i = 1, \dots, n - 1, \quad |\mathcal{P}'(\mathbf{g}_n)|^2 = \beta_n + \beta_{n+1}.$$

We shall denote by  $\bar{\mathcal{P}} : \mathbf{L} \rightarrow \mathbf{L}'$  the orthogonal projection of the space  $\mathbf{L}$  onto  $\mathbf{L}'$  (in this case, of course,  $\bar{\mathcal{P}}(\mathbf{e}) = \mathbf{0}$ ), and we construct in  $\mathbf{L}$  an orthonormal basis  $\bar{\mathbf{g}}_1, \dots, \bar{\mathbf{g}}_{n+1}$  for which  $|\bar{\mathcal{P}}(\bar{\mathbf{g}}_i)|^2 = \beta_i$  for all  $i = 1, \dots, n + 1$ .

Let us set  $\bar{\mathbf{g}}_i = \mathbf{g}_i$  for  $i = 1, \dots, n - 2$  and  $\bar{\mathbf{g}}_n = a\mathbf{g}_n + b\mathbf{e}$ ,  $\bar{\mathbf{g}}_{n+1} = c\mathbf{g}_n + d\mathbf{e}$ , where the numbers  $a, b, c, d$  are chosen in such a way that the following conditions are satisfied:

$$\begin{aligned} |\bar{\mathbf{g}}_n| &= |\bar{\mathbf{g}}_{n+1}| = 1, & (\bar{\mathbf{g}}_n, \bar{\mathbf{g}}_{n+1}) &= 0, \\ |\bar{\mathcal{P}}(\bar{\mathbf{g}}_n)|^2 &= \beta_n, & |\bar{\mathcal{P}}(\bar{\mathbf{g}}_{n+1})|^2 &= \beta_{n+1}. \end{aligned} \tag{7.45}$$

Then the system of vectors  $\bar{\mathbf{g}}_1, \dots, \bar{\mathbf{g}}_{n+1}$  proves the assertion  $P'(m, n+1)$ .

The relationships (7.45) can be rewritten in the form

$$\begin{aligned} a^2 + b^2 &= c^2 + d^2 = 1, & ac + bd &= 0, \\ a^2(\beta_n + \beta_{n+1}) &= \beta_n, & c^2(\beta_n + \beta_{n+1}) &= \beta_{n+1}. \end{aligned}$$

It is easily verified that these relationships will be satisfied if we set

$$b = \pm c, \quad d = \mp a, \quad a = \sqrt{\frac{\beta_n}{\beta_n + \beta_{n+1}}}, \quad c = \sqrt{\frac{\beta_{n+1}}{\beta_n + \beta_{n+1}}}.$$

Before proceeding to part 2, let us make the following observation.

**Proposition 7.32** *To prove the assertion  $P'(m, n)$ , we may assume that  $\beta_i < 1$  for all  $i = 1, \dots, n$ .*

*Proof* Let  $1 = \beta_1 = \dots = \beta_k > \beta_{k+1} \geq \dots \geq \beta_n > 0$ . We choose in the  $n$ -dimensional vector space  $\mathbb{L}$  an arbitrary subspace  $\mathbb{L}_k$  of dimension  $k$  and consider the orthogonal decomposition  $\mathbb{L} = \mathbb{L}_k \oplus \mathbb{L}_k^\perp$ . We note that

$$1 > \beta_{k+1} \geq \dots \geq \beta_n > 0 \quad \text{and} \quad \beta_{k+1} + \dots + \beta_n = m - k.$$

Therefore, if the assertion  $P'(m - k, n - k)$  holds for the numbers  $\beta_{k+1}, \dots, \beta_n$ , then in  $\mathbb{L}_k^\perp$ , there exist a subspace  $\mathbb{L}'_k$  of dimension  $m - k$  and an orthonormal basis  $\mathbf{g}_{k+1}, \dots, \mathbf{g}_n$  such that  $|\mathcal{P}(\mathbf{g}_i)|^2 = \beta_i$  for  $i = k + 1, \dots, n$ , where  $\mathcal{P} : \mathbb{L}_k^\perp \rightarrow \mathbb{L}'_k$  is an orthogonal projection.

We now set  $\mathbb{L}' = \mathbb{L}_k \oplus \mathbb{L}'_k$  and choose in  $\mathbb{L}_k$  an arbitrary orthonormal basis  $\mathbf{g}_1, \dots, \mathbf{g}_k$ . Then if  $\mathcal{P}' : \mathbb{L} \rightarrow \mathbb{L}'$  is the orthogonal projection, we have that  $|\mathcal{P}'(\mathbf{g}_i)|^2 = 1$  for  $i = 1, \dots, k$  and  $|\mathcal{P}'(\mathbf{g}_i)|^2 = \beta_i$  for  $i = k + 1, \dots, n$ .  $\square$

*Part 2:* Assertion  $P'(m, n)$  implies assertion  $P'(n, m - n)$ . Let us consider  $n$  numbers  $\beta_1 \geq \dots \geq \beta_n$  satisfying condition (7.44) in which the number  $m$  is replaced by  $n - m$ . We must construct an orthogonal projection  $\mathcal{P}' : \mathbb{L} \rightarrow \mathbb{L}'$  of the  $n$ -dimensional Euclidean space  $\mathbb{L}$  onto the  $(m - n)$ -dimensional subspace  $\mathbb{L}'$  and an orthonormal basis  $\mathbf{g}_1, \dots, \mathbf{g}_n$  in  $\mathbb{L}$  for which the conditions  $|\mathcal{P}'(\mathbf{g}_i)|^2 = \beta_i$ ,  $i = 1, \dots, n$ , are satisfied. By a previous observation, we may assume that all  $\beta_i$  are less than 1. Then the numbers  $\beta'_i = 1 - \beta_i$  satisfy conditions (7.44), and by assertion  $P'(m, n)$ , there exist an orthonormal projection  $\bar{\mathcal{P}} : \mathbb{L} \rightarrow \bar{\mathbb{L}}$  of the space  $\mathbb{L}$  onto the  $m$ -dimensional subspace  $\bar{\mathbb{L}}$  and an orthonormal basis  $\mathbf{g}_1, \dots, \mathbf{g}_n$  for which the conditions  $|\bar{\mathcal{P}}(\mathbf{g}_i)|^2 = \beta'_i$  are satisfied. For the desired  $(m - n)$ -dimensional subspace we shall take  $\mathbb{L}' = \bar{\mathbb{L}}^\perp$  and denote by  $\mathcal{P}'$  the orthogonal projection onto  $\mathbb{L}'$ . Then for each  $i = 1, \dots, n$ , the equalities

$$\mathbf{g}_i = \bar{\mathcal{P}}(\mathbf{g}_i) + \mathcal{P}'(\mathbf{g}_i), \quad 1 = |\mathbf{g}_i|^2 = |\bar{\mathcal{P}}(\mathbf{g}_i)|^2 + |\mathcal{P}'(\mathbf{g}_i)|^2 = \beta'_i + |\mathcal{P}'(\mathbf{g}_i)|^2$$

are satisfied, from which it follows that  $|\mathcal{P}'(\mathbf{g}_i)|^2 = 1 - \beta'_i = \beta_i$ .

*Part 3:* Assertion  $P'(m+1, n)$  for all  $n > m+1$  is a consequence of  $P'(m', n)$  for all  $m' \leq m$  and  $n > m'$ . By our assumption, the assertion  $P'(m, n)$  holds in particular for  $n = 2m+1$ . By part 2, we may assert that  $P'(m+1, 2m+1)$  holds, and since  $2(m+1) \leq (2m+1) + 1$ , then by virtue of part 1, we may conclude that  $P'(m+1, n)$  holds for all  $n \geq 2m+1$ . It remains to prove the assertions  $P'(m+1, n)$  for  $m+2 \leq n \leq 2m$ . But these assertions follow from  $P'(n-(m+1), n)$  by part 2. It is necessary only to verify that the inequalities  $1 \leq n-(m+1) \leq m$  are satisfied, which follows directly from the assumption that  $m+2 \leq n \leq 2m$ .

## 7.5 Symmetric Transformations

As we observed at the beginning of Sect. 7.1, for a Euclidean space  $L$ , there exists a natural isomorphism  $L \xrightarrow{\sim} L^*$  that allows us to identify in this case the space  $L^*$  with  $L$ . In particular, using the definition given in Sect. 3.7, we may define for an arbitrary basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the space  $L$  the *dual basis*  $\mathbf{f}_1, \dots, \mathbf{f}_n$  of the space  $L$  by the condition  $(\mathbf{f}_i, \mathbf{e}_i) = 1, (\mathbf{f}_i, \mathbf{e}_j) = 0$  for  $i \neq j$ . Thus an orthonormal basis is one that is its own dual.

In the same way, we can assume that for an arbitrary linear transformation  $\mathcal{A} : L \rightarrow L$ , the dual transformation  $\mathcal{A}^* : L^* \rightarrow L^*$  defined in Sect. 3.7 is a linear transformation of the Euclidean space  $L$  into itself and is determined by the condition

$$(\mathcal{A}^*(\mathbf{x}), \mathbf{y}) = (\mathbf{x}, \mathcal{A}(\mathbf{y})) \quad (7.46)$$

for all vectors  $\mathbf{x}, \mathbf{y} \in L$ . By Theorem 3.81, the matrix of the linear transformation  $\mathcal{A}$  in an arbitrary basis of the space  $L$  and the matrix of the dual transformation  $\mathcal{A}^*$  in the dual basis are transposes of each other. In particular, the matrices of the transformations  $\mathcal{A}$  and  $\mathcal{A}^*$  in an arbitrary orthonormal basis are transposes of each other. This is in accord with the notation  $A^*$  that we have chosen for the transpose matrix. It is easily verified also that conversely, if the matrices of transformations  $\mathcal{A}$  and  $\mathcal{B}$  in some orthonormal basis are transposes of each other, then the transformations  $\mathcal{A}$  and  $\mathcal{B}$  are dual.

As an example, let us consider the orthogonal transformation  $\mathcal{U}$ , for which by definition, the condition  $(\mathcal{U}(\mathbf{x}), \mathcal{U}(\mathbf{y})) = (\mathbf{x}, \mathbf{y})$  is satisfied. By formula (7.46), we have the equality  $(\mathcal{U}(\mathbf{x}), \mathcal{U}(\mathbf{y})) = (\mathbf{x}, \mathcal{U}^*\mathcal{U}(\mathbf{y}))$ , from which follows  $(\mathbf{x}, \mathcal{U}^*\mathcal{U}(\mathbf{y})) = (\mathbf{x}, \mathbf{y})$ . This implies that  $(\mathbf{x}, \mathcal{U}^*\mathcal{U}(\mathbf{y}) - \mathbf{y}) = 0$  for all vectors  $\mathbf{x}$ , from which follows the equality  $\mathcal{U}^*\mathcal{U}(\mathbf{y}) = \mathbf{y}$  for all vectors  $\mathbf{y} \in L$ . In other words, the fact that  $\mathcal{U}^*\mathcal{U}$  is equal to  $\mathcal{E}$ , the identity transformation, is equivalent to the property of orthogonality of the transformation  $\mathcal{U}$ . In matrix form, this is the relationship (7.18).

**Definition 7.33** A linear transformation  $\mathcal{A}$  of a Euclidean space is called *symmetric* or *self-dual* if  $\mathcal{A}^* = \mathcal{A}$ .

In other words, for a symmetric transformation  $\mathcal{A}$  and arbitrary vectors  $\mathbf{x}$  and  $\mathbf{y}$ , the following condition must be satisfied:

$$(\mathcal{A}(\mathbf{x}), \mathbf{y}) = (\mathbf{x}, \mathcal{A}(\mathbf{y})), \quad (7.47)$$

that is, the bilinear form  $\varphi(\mathbf{x}, \mathbf{y}) = (\mathcal{A}(\mathbf{x}), \mathbf{y})$  is symmetric. As we have seen, from this it follows that in an arbitrary orthonormal basis, the matrix of the transformation  $\mathcal{A}$  is symmetric.

Symmetric linear transformations play a very large role in mathematics and its applications. Their most essential applications relate to quantum mechanics, where symmetric transformations of infinite-dimensional Hilbert space (see the note on p. 214) correspond to what are called *observed* physical quantities. We shall, however, restrict our attention to finite-dimensional spaces. As we shall see in the sequel, even with this restriction, the theory of symmetric linear transformations has a great number of applications.

The following theorem gives a basic property of symmetric linear transformations of finite-dimensional Euclidean spaces.

**Theorem 7.34** *Every symmetric linear transformation of a real vector space has an eigenvector.*

In view of the very large number of applications of this theorem, we shall present three proofs, based on different principles.

*Proof of Theorem 7.34 First proof.* Let  $\mathcal{A}$  be a symmetric linear transformation of a Euclidean space  $L$ . If  $\dim L > 2$ , then by Theorem 4.22, it has a one- or two-dimensional invariant subspace  $L'$ . It is obvious that the restriction of the transformation  $\mathcal{A}$  to the invariant subspace  $L'$  is also a symmetric transformation. If  $\dim L' = 1$ , then we have  $L' = \langle \mathbf{e} \rangle$ , where  $\mathbf{e} \neq \mathbf{0}$ , and this implies that  $\mathbf{e}$  is an eigenvector. Consequently, to prove the theorem, it suffices to show that a symmetric linear transformation in the two-dimensional subspace  $L'$  has an eigenvector. Choosing in  $L'$  an orthonormal basis, we obtain for  $\mathcal{A}$  a symmetric matrix in this basis:

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

In order to find an eigenvector of the transformation  $\mathcal{A}$ , we must find a *real* root of the polynomial  $|A - tE|$ . This polynomial has the form

$$(a - t)(c - t) - b^2 = t^2 - (a + c)t + ac - b^2$$

and has a real root if and only if its discriminant is nonnegative. But the discriminant of this quadratic trinomial is equal to

$$(a + c)^2 - 4(ac - b^2) = (a - c)^2 + 4b^2 \geq 0,$$

and the proof is complete.

**Second proof.** The second proof is based on the complexification  $L^{\mathbb{C}}$  of the real vector space  $L$ . Following the construction presented in Sect. 4.3, we may extend the transformation  $\mathcal{A}$  to the vectors of the space  $L^{\mathbb{C}}$ . By Theorem 4.18, the obtained transformation  $\mathcal{A}^{\mathbb{C}} : L^{\mathbb{C}} \rightarrow L^{\mathbb{C}}$  will already have an eigenvector  $e \in L^{\mathbb{C}}$  and eigenvalue  $\lambda \in \mathbb{C}$ , so that  $\mathcal{A}^{\mathbb{C}}(e) = \lambda e$ .

We shall extend the inner product  $(x, y)$  from the space  $L$  to  $L^{\mathbb{C}}$  so that it determines there a Hermitian form (see the definition on p. 210). It is clear that this can be accomplished in only one way: defining two vectors  $a_1 = x_1 + i y_1$  and  $a_2 = x_2 + i y_2$  of the space  $L^{\mathbb{C}}$ , we obtain the inner product according to the formula

$$(a_1, a_2) = (x_1, x_2) + (y_1, y_2) + i((y_1, x_2) - (x_1, y_2)). \quad (7.48)$$

The verification of the fact that the inner product  $(a_1, a_2)$  thus defined actually determines in  $L^{\mathbb{C}}$  a Hermitian form is reduced to the verification of sesquilinearity (in this case, it suffices to consider separately the product of a vector  $a_1$  and a vector  $a_2$  by a real number and by  $i$ ) and the property of being Hermitian. Here all calculations are completely trivial, and we shall omit them.

An important new property of the inner product  $(a_1, a_2)$  that we have obtained is its positive definiteness, that is, like the scalar product  $(a, a)$ , it is real (this follows from the Hermitian property) and  $(a, a) > 0$ ,  $a \neq 0$  (this is a direct consequence of formula (7.48), for  $x_1 = x_2$ ,  $y_1 = y_2$ ). It is obvious that for the new inner product we also have an analogue of the relationship (7.47), that is,

$$(\mathcal{A}^{\mathbb{C}}(a_1), a_2) = \overline{(a_1, \mathcal{A}^{\mathbb{C}}(a_2))}; \quad (7.49)$$

in other words, the form  $\varphi(a_1, a_2) = (\mathcal{A}^{\mathbb{C}}(a_1), a_2)$  is Hermitian. Let us apply (7.49) to the vectors  $a_1 = a_2 = e$ . Then we obtain  $(\lambda e, e) = (e, \lambda e)$ . Taking into account the Hermitian property, we have the equalities  $(\lambda e, e) = \lambda(e, e)$  and  $(e, \lambda e) = \overline{\lambda}(e, e)$ , from which it follows that  $\lambda(e, e) = \overline{\lambda}(e, e)$ . Since  $(e, e) > 0$ , we derive from this that  $\lambda = \overline{\lambda}$ , that is, the number  $\lambda$  is real. Thus the characteristic polynomial  $|\mathcal{A}^{\mathbb{C}} - t\mathcal{E}|$  of the transformation  $\mathcal{A}^{\mathbb{C}}$  has a real root  $\lambda$ . But a basis of the space  $L$  as a space over  $\mathbb{R}$  is a basis of the space  $L^{\mathbb{C}}$  over  $\mathbb{C}$ , and the matrix of the transformation  $\mathcal{A}^{\mathbb{C}}$  in this basis coincides with the matrix of the transformation  $\mathcal{A}$ . In other words,  $|\mathcal{A}^{\mathbb{C}} - t\mathcal{E}| = |\mathcal{A} - t\mathcal{E}|$ , which implies that the characteristic polynomial  $|\mathcal{A} - t\mathcal{E}|$  of the transformation  $\mathcal{A}$  has a real root  $\lambda$ , and this implies that the transformation  $\mathcal{A} : L \rightarrow L$  has an eigenvector in the space  $L$ .

**Third proof.** The third proof rests on certain facts from analysis, which we now introduce. We first observe that a Euclidean space can be naturally converted into a metric space by defining the distance  $r(x, y)$  between two vectors  $x$  and  $y$  by the relationship  $r(x, y) = |x - y|$ . Thus in the Euclidean space  $L$  we have the notions of convergence, limit, continuous functions, and closed and bounded sets; see p. xvii.

The *Bolzano–Weierstrass theorem* asserts that for an arbitrary closed and bounded set  $X$  in a finite-dimensional Euclidean space  $L$  and arbitrary continuous function  $\varphi(x)$  on  $X$  there exists a vector  $x_0 \in X$  at which  $\varphi(x)$  assumes its

maximum value: that is,  $\varphi(\mathbf{x}_0) \geq \varphi(\mathbf{x})$  for all  $\mathbf{x} \in X$ . This theorem is well known from real analysis in the case that the set  $X$  is an interval of the real line. Its proof in the general case is exactly the same and is usually presented somewhat later. Here we shall use the theorem without offering a proof.

Let us apply the Bolzano–Weierstrass theorem to the set  $X$  consisting of all vectors  $\mathbf{x}$  of the space  $L$  such that  $|\mathbf{x}| = 1$ , that is, to the sphere of radius 1, and to the function  $\varphi(\mathbf{x}) = (\mathbf{x}, \mathcal{A}(\mathbf{x}))$ . This function is continuous not only on  $X$ , but also on the entire space  $L$ . Indeed, it suffices to choose in the space  $L$  an arbitrary basis and to write down in it the inner product  $(\mathbf{x}, \mathcal{A}(\mathbf{x}))$  as a quadratic form in the coordinates of the vector  $\mathbf{x}$ . Of importance to us is solely the fact that as a result, we obtain a *polynomial* in the coordinates. After this, it suffices to use the well-known theorem that states that the sum and product of continuous functions are continuous. Then the question is reduced to a verification of the fact that an arbitrary coordinate of the vector  $\mathbf{x}$  is a continuous function of  $\mathbf{x}$ , but this is completely obvious.

Thus the function  $(\mathbf{x}, \mathcal{A}(\mathbf{x}))$  assumes its maximum over the set  $X$  at some  $\mathbf{x}_0 = \mathbf{e}$ . Let us denote this value by  $\lambda$ . Consequently,  $(\mathbf{x}, \mathcal{A}(\mathbf{x})) \leq \lambda$  for every  $\mathbf{x}$  for which  $|\mathbf{x}| = 1$ . For every nonnull vector  $\mathbf{y}$ , we set  $\mathbf{x} = \mathbf{y}/|\mathbf{y}|$ . Then  $|\mathbf{x}| = 1$ , and applying to this vector the inequality above, we see that  $(\mathbf{y}, \mathcal{A}(\mathbf{y})) \leq \lambda(\mathbf{y}, \mathbf{y})$  for all  $\mathbf{y}$  (this obviously holds as well for  $\mathbf{y} = \mathbf{0}$ ).

Let us prove that the number  $\lambda$  is an eigenvalue of the transformation  $\mathcal{A}$ . To this end, let us write the condition that defines  $\lambda$  in the form

$$(\mathbf{y}, \mathcal{A}(\mathbf{y})) \leq \lambda(\mathbf{y}, \mathbf{y}), \quad \lambda = (\mathbf{e}, \mathcal{A}(\mathbf{e})), \quad |\mathbf{e}| = 1, \quad (7.50)$$

for an arbitrary vector  $\mathbf{y} \in L$ .

Let us apply (7.50) to the vector  $\mathbf{y} = \mathbf{e} + \varepsilon\mathbf{z}$ , where both the scalar  $\varepsilon$  and vector  $\mathbf{z} \in L$  are thus far arbitrary. Expanding the expressions  $(\mathbf{y}, \mathcal{A}(\mathbf{y})) = (\mathbf{e} + \varepsilon\mathbf{z}, \mathcal{A}(\mathbf{e} + \varepsilon\mathcal{A}(\mathbf{z})))$  and  $(\mathbf{y}, \mathbf{y}) = (\mathbf{e} + \varepsilon\mathbf{z}, \mathbf{e} + \varepsilon\mathbf{z})$ , we obtain the inequality

$$\begin{aligned} & (\mathbf{e}, \mathcal{A}(\mathbf{e})) + \varepsilon(\mathbf{e}, \mathcal{A}(\mathbf{z})) + \varepsilon(\mathbf{z}, \mathcal{A}(\mathbf{e})) + \varepsilon^2(\mathcal{A}(\mathbf{z}), \mathcal{A}(\mathbf{z})) \\ & \leq \lambda((\mathbf{e}, \mathbf{e}) + \varepsilon(\mathbf{e}, \mathbf{z}) + \varepsilon(\mathbf{z}, \mathbf{e}) + \varepsilon^2(\mathbf{z}, \mathbf{z})). \end{aligned}$$

In view of the symmetry of the transformation  $\mathcal{A}$ , on the basis of the properties of Euclidean spaces and recalling that  $(\mathbf{e}, \mathbf{e}) = 1$ ,  $(\mathbf{e}, \mathcal{A}(\mathbf{e})) = \lambda$ , after canceling the common term  $(\mathbf{e}, \mathcal{A}(\mathbf{e})) = \lambda(\mathbf{e}, \mathbf{e})$  on both sides of the above inequality, we obtain

$$2\varepsilon(\mathbf{e}, \mathcal{A}(\mathbf{z}) - \lambda\mathbf{z}) + \varepsilon^2((\mathcal{A}(\mathbf{z}), \mathcal{A}(\mathbf{z})) - \lambda(\mathbf{z}, \mathbf{z})) \leq 0. \quad (7.51)$$

Let us now note that every expression  $a\varepsilon + b\varepsilon^2$  in the case  $a \neq 0$  assumes a positive value for some  $\varepsilon$ . For this it is necessary to choose a value  $|\varepsilon|$  sufficiently small that  $a + b\varepsilon$  has the same sign as  $a$ , and then to choose the appropriate sign for  $\varepsilon$ . Thus the inequality (7.51) always leads to a contradiction except in the case  $(\mathbf{e}, \mathcal{A}(\mathbf{z}) - \lambda\mathbf{z}) = 0$ .

If for some vector  $\mathbf{z} \neq \mathbf{0}$ , we have  $\mathcal{A}(\mathbf{z}) = \lambda\mathbf{z}$ , then  $\mathbf{z}$  is an eigenvector of the transformation  $\mathcal{A}$  with eigenvalue  $\lambda$ , which is what we wished to prove. But if



$\mathcal{A}(z) - \lambda z \neq \mathbf{0}$  for all  $z \neq \mathbf{0}$ , then the kernel of the transformation  $\mathcal{A} - \lambda\mathcal{E}$  is equal to  $\{\mathbf{0}\}$ . From Theorem 3.68 it follows that then the transformation  $\mathcal{A} - \lambda\mathcal{E}$  is an isomorphism, and its image is equal to all of the space  $L$ . This implies that for arbitrary  $u \in L$ , it is possible to choose a vector  $z \in L$  such that  $u = \mathcal{A}(z) - \lambda z$ . Then taking into account the relationship  $(e, \mathcal{A}(z) - \lambda z) = 0$ , we obtain that an arbitrary vector  $u \in L$  satisfies the equality  $(e, u) = 0$ . But this is impossible at least for  $u = e$ , since  $|e| = 1$ .  $\square$

The further theory of symmetric transformations is constructed on the basis of some very simple considerations.

**Theorem 7.35** *If a subspace  $L'$  of a Euclidean space  $L$  is invariant with respect to the symmetric transformation  $\mathcal{A}$ , then its orthogonal complement  $(L')^\perp$  is also invariant.*

*Proof* The result is a direct consequence of the definitions. Let  $y$  be a vector in  $(L')^\perp$ . Then  $(x, y) = 0$  for all  $x \in L'$ . In view of the symmetry of the transformation  $\mathcal{A}$ , we have the relationship

$$(x, \mathcal{A}(y)) = (\mathcal{A}(x), y),$$

while taking into account the invariance of  $L'$  yields that  $\mathcal{A}(x) \in L'$ . This implies that  $(x, \mathcal{A}(y)) = 0$  for all vectors  $x \in L'$ , that is,  $\mathcal{A}(y) \in (L')^\perp$ , and this completes the proof of the theorem.  $\square$

Combining Theorems 7.34 and 7.35 yields a fundamental result in the theory of symmetric transformations.

**Theorem 7.36** *For every symmetric transformation  $\mathcal{A}$  of a Euclidean space  $L$  of finite dimension, there exists an orthonormal basis of this space consisting of eigenvectors of the transformation  $\mathcal{A}$ .*

*Proof* The proof is by induction on the dimension of the space  $L$ . Indeed, by Theorem 7.34, the transformation  $\mathcal{A}$  has at least one eigenvector  $e$ . Let us set

$$L = \langle e \rangle \oplus \langle e \rangle^\perp,$$

where  $\langle e \rangle^\perp$  has dimension  $n - 1$ , and by Theorem 7.35, is invariant with respect to  $\mathcal{A}$ . By the induction hypothesis, in the space  $\langle e \rangle^\perp$  there exists a required basis. If we add the vector  $e$  to this basis, we obtain the desired basis in  $L$ .  $\square$

Let us discuss this result. For a symmetric transformation  $\mathcal{A}$ , we have an orthonormal basis  $e_1, \dots, e_n$  consisting of eigenvectors. But to what extent is such a basis uniquely determined? Suppose the vector  $e_i$  has the associated eigenvalue  $\lambda_i$ .

Then in our basis, the transformation  $\mathcal{A}$  has matrix

$$A = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}. \quad (7.52)$$

But as we saw in Sect. 4.1, the eigenvalues of a linear transformation  $\mathcal{A}$  coincide with the roots of the characteristic polynomial

$$|\mathcal{A} - t\mathcal{E}| = |A - tE| = \prod_{i=1}^n (\lambda_i - t).$$

Thus the eigenvalues  $\lambda_1, \dots, \lambda_n$  of the transformation  $\mathcal{A}$  are uniquely determined. Suppose that the distinct values among them are  $\lambda_1, \dots, \lambda_k$ . If we assemble all the vectors of the constructed orthonormal basis that correspond to one and the same eigenvalue  $\lambda_i$  (from the set  $\lambda_1, \dots, \lambda_k$  of distinct eigenvalues) and consider the subspace spanned by them, then we obviously obtain the eigensubspace  $L_{\lambda_i}$  (see the definition on p. 138). We then have the orthogonal decomposition

$$L = L_{\lambda_1} \oplus \cdots \oplus L_{\lambda_k}, \quad \text{where } L_{\lambda_i} \perp L_{\lambda_j} \text{ for all } i \neq j. \quad (7.53)$$

The restriction of  $\mathcal{A}$  to the eigensubspace  $L_{\lambda_i}$  gives a transformation  $\lambda_i \mathcal{E}$ , and in this subspace, every orthonormal basis consists of eigenvectors (with eigenvalue  $\lambda_i$ ).

Thus we see that a symmetric transformation  $\mathcal{A}$  uniquely defines only the eigensubspace  $L_{\lambda_i}$ , while in each of them, one can choose an orthonormal basis as one likes. On combining these bases, we obtain an arbitrary basis of the space  $L$  satisfying the conditions of Theorem 7.36.

Let us note that every eigenvector of the transformation  $\mathcal{A}$  lies in one of the subspaces  $L_{\lambda_i}$ . If two eigenvectors  $\mathbf{x}$  and  $\mathbf{y}$  are associated with different eigenvalues  $\lambda_i \neq \lambda_j$ , then they lie in different subspaces  $L_{\lambda_i}$  and  $L_{\lambda_j}$ , and in view of the orthogonality of the decomposition (7.53), they must be orthogonal. We thus obtain the following result.

**Theorem 7.37** *The eigenvectors of a symmetric transformation corresponding to different eigenvalues are orthogonal.*

We note that this theorem can also be easily proved by direct calculation.

*Proof of Theorem 7.37* Let  $\mathbf{x}$  and  $\mathbf{y}$  be eigenvectors of a symmetric transformation  $\mathcal{A}$  corresponding to distinct eigenvalues  $\lambda_i$  and  $\lambda_j$ . Let us substitute the expressions  $\mathcal{A}(\mathbf{x}) = \lambda_i \mathbf{x}$  and  $\mathcal{A}(\mathbf{y}) = \lambda_j \mathbf{y}$  into the equality  $(\mathcal{A}(\mathbf{x}), \mathbf{y}) = (\mathbf{x}, \mathcal{A}(\mathbf{y}))$ . From this we obtain  $(\lambda_i - \lambda_j)(\mathbf{x}, \mathbf{y}) = 0$ , and since  $\lambda_i \neq \lambda_j$ , we have  $(\mathbf{x}, \mathbf{y}) = 0$ .  $\square$

Theorem 7.36 is often formulated conveniently as a theorem about quadratic forms using Theorem 6.3 from Sect. 6.1 and the possibility of identifying the space

$L^*$  with  $L$  if the space  $L$  is equipped with an inner product. Indeed, Theorem 6.3 shows that every bilinear form  $\varphi$  on a Euclidean space  $L$  can be represented in the form

$$\varphi(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathcal{A}(\mathbf{y})), \quad (7.54)$$

where  $\mathcal{A}$  is the linear transformation of the space  $L$  to  $L^*$  uniquely defined by the bilinear form  $\varphi$ ; that is, if we make the identification of  $L^*$  with  $L$ , it is a transformation of the space  $L$  into itself.

It is obvious that the symmetry of the transformation  $\mathcal{A}$  coincides with the symmetry of the bilinear form  $\varphi$ . Therefore, the bijection between symmetric bilinear forms and linear transformations established above yields the same correspondence between quadratic forms and symmetric linear transformations of a Euclidean space  $L$ . Moreover, in view of relationship (7.54), to the symmetric transformation  $\mathcal{A}$  there corresponds the quadratic form

$$\psi(\mathbf{x}) = (\mathbf{x}, \mathcal{A}(\mathbf{x})),$$

and every quadratic form  $\psi(\mathbf{x})$  has a unique representation in this form.

If in some basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , the transformation  $\mathcal{A}$  has a diagonal matrix (7.52), then for the vector  $\mathbf{x} = x_1\mathbf{e}_1 + \dots + x_n\mathbf{e}_n$ , the quadratic form  $\psi(\mathbf{x})$  has in this basis the canonical form

$$\psi(\mathbf{x}) = \lambda_1 x_1^2 + \dots + \lambda_n x_n^2. \quad (7.55)$$

Thus Theorem 7.36 is equivalent to the following.

**Theorem 7.38** *For any quadratic form in a finite-dimensional Euclidean space, there exists an orthonormal basis in which it has the canonical form (7.55).*

Theorem 7.38 is sometimes conveniently formulated as a theorem about arbitrary vector spaces.

**Theorem 7.39** *For two quadratic forms in a finite-dimensional vector space, one of which is positive definite, there exists a basis (not necessarily orthonormal) in which they both have canonical form (7.55).*

In this case, we say that in a suitable basis, these quadratic forms are reduced to a sum of squares (even if there are negative coefficients  $\lambda_i$  in formula (7.55)).

*Proof of Theorem 7.39* Let  $\psi_1(\mathbf{x})$  and  $\psi_2(\mathbf{x})$  be two such quadratic forms, one of which, let it be  $\psi_1(\mathbf{x})$ , is positive definite. By Theorem 6.10, there exists, in the vector space  $L$  in question, a basis in which the form  $\psi_1(\mathbf{x})$  has the canonical form (7.55). Since by assumption, the quadratic form  $\psi_1(\mathbf{x})$  is positive definite, it follows that in formula (7.55), all the numbers  $\lambda_i$  are positive, and therefore, there exists a basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the space  $L$  in which  $\psi_1(\mathbf{x})$  is brought into the form

$$\psi(\mathbf{x}) = x_1^2 + \dots + x_n^2. \quad (7.56)$$

Let us consider as the scalar product  $(x, y)$  in the space  $L$  the symmetric bilinear form  $\varphi(x, y)$ , associated by Theorem 6.6 with the quadratic form  $\psi_1(x)$ . We thereby convert  $L$  into a Euclidean space.

As can be seen from formulas (6.14) and (7.56), the basis  $e_1, \dots, e_n$  for this inner product is orthonormal. Then by Theorem 7.38, there exists an orthonormal basis  $e'_1, \dots, e'_n$  of the space  $L$  in which the form  $\psi_2(x)$  has canonical form (7.55). But since the basis  $e'_1, \dots, e'_n$  is orthonormal with respect to the inner product that we defined with the help of the quadratic form  $\psi_1(x)$ , then in this basis,  $\psi_1(x)$  as before takes the form (7.56), and that completes the proof of the theorem.  $\square$

*Remark 7.40* It is obvious that Theorem 7.39 remains true if in its formulation we replace the condition of *positive* definiteness of one of the forms by the condition of *negative* definiteness. Indeed, if  $\psi(x)$  is a negative definite quadratic form, then the form  $-\psi(x)$  is positive definite, and both of these assume canonical form in one and the same basis.

Without the assumption of positive (or negative) definiteness of one of the quadratic forms, Theorem 7.39 is no longer true. To prove this, let us derive one *necessary* (but not sufficient) condition for two quadratic forms  $\psi_1(x)$  and  $\psi_2(x)$  to be simultaneously reduced to a sum of squares. Let  $A_1$  and  $A_2$  be their matrices in some basis. If the quadratic forms  $\psi_1(x)$  and  $\psi_2(x)$  are simultaneously reducible to sums of squares, then in some other basis, their matrices  $A'_1$  and  $A'_2$  will be diagonal, that is,

$$A'_1 = \begin{pmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_n \end{pmatrix}, \quad A'_2 = \begin{pmatrix} \beta_1 & 0 & \cdots & 0 \\ 0 & \beta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \beta_n \end{pmatrix}.$$

Then the polynomial  $|A'_1 t + A'_2|$  is equal to  $\prod_{i=1}^n (\alpha_i t + \beta_i)$ , that is, it can be factored as a product of linear factors  $\alpha_i t + \beta_i$ . But by formula (6.10) for replacing the matrix of a bilinear form through a change of basis, the matrices  $A_1, A'_1$  and  $A_2, A'_2$  are related by

$$A'_1 = C^* A_1 C, \quad A'_2 = C^* A_2 C,$$

where  $C$  is some nonsingular matrix, that is,  $|C| \neq 0$ . Therefore,

$$|A'_1 t + A'_2| = |C^* (A_1 t + A_2) C| = |C^*| |A_1 t + A_2| |C|,$$

from which taking into account the equality  $|C^*| = |C|$ , we obtain the relationship

$$|A_1 t + A_2| = |C|^{-2} |A'_1 t + A'_2|,$$

from which it follows that the polynomial  $|A_1 t + A_2|$  can also be factored into linear factors. Thus for two quadratic forms  $\psi_1(x)$  and  $\psi_2(x)$  with matrices  $A_1$  and  $A_2$  to be simultaneously reduced each to a sum of squares, it is *necessary* that the polynomial  $|A_1 t + A_2|$  be factorable into real linear factors.

Now for  $n = 2$  we set  $\psi_1(\mathbf{x}) = x_1^2 - x_2^2$  and  $\psi_2(\mathbf{x}) = x_1x_2$ . These quadratic forms are neither positive definite nor negative definite. Their matrices have the form

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

and it is obvious that the polynomial  $|A_1t + A_2| = -(t^2 + 1)$  cannot be factored into real linear factors. This implies that the quadratic forms  $\psi_1(\mathbf{x})$  and  $\psi_2(\mathbf{x})$  cannot simultaneously be reduced to sums of squares.

The question of reducing pairs of quadratic forms with complex coefficients to sums of squares (with the help of a complex linear transformation) is examined in detail, for instance, in the book *The Theory of Matrices*, by F.R. Gantmacher. See the references section.

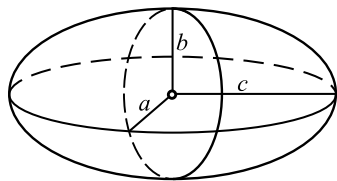
*Remark 7.41* The last proof of Theorem 7.34 that we gave makes it possible to interpret the largest eigenvalue  $\lambda$  of a symmetric transformation  $\mathcal{A}$  as the maximum of the quadratic form  $(\mathbf{x}, \mathcal{A}(\mathbf{x}))$  on the sphere  $|\mathbf{x}| = 1$ . Let  $\lambda_i$  be the other eigenvalues, so that  $(\mathbf{x}, \mathcal{A}(\mathbf{x})) = \lambda_1x_1^2 + \cdots + \lambda_nx_n^2$ . Then  $\lambda$  is the greatest among the  $\lambda_i$ . Indeed, let us assume that the eigenvalues are numbered in descending order:  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ . Then

$$\lambda_1x_1^2 + \cdots + \lambda_nx_n^2 \leq \lambda_1(x_1^2 + \cdots + x_n^2),$$

and the maximum value of the form  $(\mathbf{x}, \mathcal{A}(\mathbf{x}))$  on the sphere  $|\mathbf{x}| = 1$  is equal to  $\lambda_1$  (it is attained at the vector with coordinates  $x_1 = 1, x_2 = \cdots = x_n = 0$ ). This implies that  $\lambda_1 = \lambda$ .

There is an analogous characteristic for the other eigenvalues  $\lambda_i$  as well, namely the *Courant–Fischer theorem*, which we shall present without proof. Let us consider all possible vector subspaces  $L' \subset L$  of dimension  $k$ . We restrict the quadratic form  $(\mathbf{x}, \mathcal{A}(\mathbf{x}))$  to the subspace  $L'$  and examine its values at the intersection of  $L'$  with the unit sphere, that is, the set of all vectors  $\mathbf{x} \in L'$  that satisfy  $|\mathbf{x}| = 1$ . By the Bolzano–Weierstrass theorem, the restriction of the form  $(\mathbf{x}, \mathcal{A}(\mathbf{x}))$  to  $L'$  assumes a maximum value  $\lambda'$  at some point of the sphere, which, of course depends on the subspace  $L'$ . The Courant–Fischer theorem asserts that the smallest number thus obtained (as the subspace  $L'$  ranges over all subspaces of dimension  $k$ ) is equal to the eigenvalue  $\lambda_{n-k+1}$ .

*Remark 7.42* Eigenvectors are connected with the question of finding maxima and minima. Let  $f(x_1, \dots, x_n)$  be a real-valued differentiable function of  $n$  real variables. A point at which all the derivatives of the function  $f$  with respect to the variables  $(x_1, \dots, x_n)$ , that is, the derivatives in all directions from this point, are equal to zero is called a *critical point* of the function. It is proved in real analysis that with some natural constraints, this condition is necessary (but not sufficient) for the function  $f$  to assume a maximum or minimum value at the point in question. Let us consider a quadratic form  $f(\mathbf{x}) = (\mathbf{x}, \mathcal{A}(\mathbf{x}))$  on the unit sphere  $|\mathbf{x}| = 1$ . It is not difficult to show that for an arbitrary point on this sphere, all points sufficiently

**Fig. 7.9** An ellipsoid

close to it can be written in some system of coordinates such that our function  $f$  can be viewed as a function of these coordinates. Then the critical points of the function  $(\mathbf{x}, \mathcal{A}(\mathbf{x}))$  are exactly those points of the sphere that are eigenvectors of the symmetric transformation  $\mathcal{A}$ .

*Example 7.43* Let an ellipsoid be given in three-dimensional space with coordinates  $x, y, z$  by the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1. \quad (7.57)$$

The expression on the left-hand side of (7.57) can be written in the form  $\psi(\mathbf{x}) = (\mathbf{x}, \mathcal{A}(\mathbf{x}))$ , where

$$\mathbf{x} = (x, y, z), \quad \mathcal{A}(\mathbf{x}) = \left( \frac{x}{a^2}, \frac{y}{b^2}, \frac{z}{c^2} \right).$$

Let us assume that  $0 < a < b < c$ . Then the maximum value that the quadratic form  $\psi(\mathbf{x})$  takes on the sphere  $|\mathbf{x}| = 1$  is  $\lambda = 1/a^2$ . It is attained on the vectors  $(\pm 1, 0, 0)$ . If  $|\psi(\mathbf{x})| \leq \lambda$  for  $|\mathbf{x}| = 1$ , then for an arbitrary vector  $\mathbf{y} \neq \mathbf{0}$ , setting  $\mathbf{x} = \mathbf{y}/|\mathbf{y}|$ , we obtain  $|\psi(\mathbf{y})| \leq \lambda|\mathbf{y}|^2$ . For the vector  $\mathbf{y} = \mathbf{0}$ , this inequality is obvious. Therefore, it holds in general for all  $\mathbf{y}$ . For  $|\psi(\mathbf{y})| = 1$ , it then follows that  $|\mathbf{y}|^2 \geq 1/\lambda$ . This implies that the shortest vector  $\mathbf{y}$  satisfying equation (7.57) is the vector  $(\pm a, 0, 0)$ . The line segments beginning at the point  $(0, 0, 0)$  and ending at the points  $(\pm a, 0, 0)$  are called the *semiminor axes* of the ellipsoid (sometimes, this same term denotes their length). Similarly, the smallest value that the quadratic form  $\psi(\mathbf{x})$  attains on the sphere  $|\mathbf{x}| = 1$  is equal to  $1/c^2$ . It attains this value at vectors  $(0, 0, \pm 1)$  on the unit sphere. Line segments corresponding to vectors  $(0, 0, \pm c)$  are called *semimajor axes* of the ellipsoid. A vector  $(0, \pm b, 0)$  corresponds to a critical point of the quadratic form  $\psi(\mathbf{x})$  that is neither a maximum nor a minimum. Such a point is called a *minimax*, that is, as it moves from this point in one direction, the function  $\psi(\mathbf{x})$  will increase, while in moving in another direction it will decrease (see Fig. 7.9). The line segments corresponding to the vectors  $(0, \pm b, 0)$  are called the *median semiaxes* of the ellipsoid.

Everything presented thus far in this chapter (with the exception of Sect. 7.3 on the orientation of a real Euclidean space) can be transferred verbatim to complex Euclidean spaces if the inner product is defined using the positive definite Hermitian form  $\varphi(\mathbf{x}, \mathbf{y})$ . The condition of positive definiteness means that for the associated quadratic Hermitian form  $\psi(\mathbf{x}) = \varphi(\mathbf{x}, \mathbf{x})$ , the inequality  $\psi(\mathbf{x}) > 0$  is satisfied for

all  $\mathbf{x} \neq \mathbf{0}$ . If we denote, as before, the inner product by  $(\mathbf{x}, \mathbf{y})$ , the last condition can be written in the form  $(\mathbf{x}, \mathbf{x}) > 0$  for all  $\mathbf{x} \neq \mathbf{0}$ .

The dual transformation  $\mathcal{A}^*$ , as previously, is defined by condition (7.46). But now, the matrix of the transformation  $\mathcal{A}^*$  in an orthonormal basis is obtained from the matrix of the transformation  $\mathcal{A}$  not simply by taking the transpose, but by taking the complex conjugate of the transpose. The analogue of a symmetric transformation is defined as a transformation  $\mathcal{A}$  whose associated bilinear form  $(\mathbf{x}, \mathcal{A}(\mathbf{y}))$  is Hermitian.

It is a fundamental fact that in quantum mechanics, one deals with *complex* space. We can formulate what was stated earlier in the following form: *observed* physical quantities correspond to Hermitian forms in infinite-dimensional complex Hilbert space.

The theory of Hermitian transformations in the finite-dimensional case is constructed even more simply than the theory of symmetric transformations in real spaces, since there is no need to prove analogues of Theorem 7.34: we know already that an arbitrary linear transformation of a complex vector space has an eigenvector. From the definition of being Hermitian, it follows that the eigenvalues of a Hermitian transformation are real. The theorems proved in this section are valid for Hermitian forms (with the same proofs).

In the complex case, a transformation  $\mathcal{U}$  preserving the inner product is called *unitary*. The reasoning carried out in Sect. 7.2 shows that for a unitary transformation  $\mathcal{U}$ , there exists an orthonormal basis consisting of eigenvectors, and all eigenvalues of the transformation  $\mathcal{U}$  are complex numbers of modulus 1.

## 7.6 Applications to Mechanics and Geometry\*

We shall present two examples from two different areas—mechanics and geometry—in which the theorems of the previous section play a key role. Since these questions will be taken up in other courses, we shall allow ourselves to be brief in both the definitions and the proofs.

*Example 7.44* Let us consider the motion of a mechanical system in a small neighborhood of its equilibrium position. One says that such a system possesses *n degrees of freedom* if in some region, its state is determined by *n* so-called *generalized coordinates*  $q_1, \dots, q_n$ , which we shall consider the coordinates of a vector  $\mathbf{q}$  in some coordinate system, and where we will take the origin  $\mathbf{0}$  to be the equilibrium position of our system. The motion of the system determines the dependence of a vector  $\mathbf{q}$  on time  $t$ . We shall assume that the equilibrium position under investigation is determined by a strict local minimum of its *potential energy*  $\Pi$ . If this value is equal to  $c$ , and the potential energy is a function  $\Pi(q_1, \dots, q_n)$  in the generalized coordinates (it is assumed that it does not depend on time), then this implies that  $\Pi(0, \dots, 0) = c$  and  $\Pi(q_1, \dots, q_n) > c$  for all remaining values  $q_1, \dots, q_n$  close to zero. From the fact that a critical point of the function  $\Pi$  corresponds to the minimum value, we may conclude that at the point  $\mathbf{0}$ , all partial derivatives  $\partial \Pi / \partial q_i$

become zero. Therefore, for an expansion of the function  $\Pi(q_1, \dots, q_n)$  as a series in powers of the variables  $q_1, \dots, q_n$  at the point  $\mathbf{0}$ , the linear terms will be equal to zero, and we obtain the expression  $\Pi(q_1, \dots, q_n) = c + \sum_{i,j=1}^n b_{ij} q_i q_j + \dots$ , where  $b_{ij}$  are certain constants, and the ellipsis indicates terms of degree greater than 2. Since we are considering motions not far from the point  $\mathbf{0}$ , we can disregard those values. It is in this approximation that we shall consider this problem. That is, we set

$$\Pi(q_1, \dots, q_n) = c + \sum_{i,j=1}^n b_{ij} q_i q_j.$$

Since  $\Pi(q_1, \dots, q_n) > c$  for all values  $q_1, \dots, q_n$  not equal to zero, the quadratic form  $\sum_{i,j=1}^n b_{ij} q_i q_j$  will be positive definite.

*Kinetic energy*  $T$  is a quadratic form in so-called *generalized velocities*  $dq_1/dt, \dots, dq_n/dt$ , which are also denoted by  $\dot{q}_1, \dots, \dot{q}_n$ , that is,

$$T = \sum_{i,j=1}^n a_{ij} \dot{q}_i \dot{q}_j, \quad (7.58)$$

where  $a_{ij} = a_{ji}$  are functions of  $\mathbf{q}$  (we assume that they do not depend on time  $t$ ). Considering as we did for potential energy only those values  $q_i$  close to zero, we may replace all the functions  $a_{ij}$  in (7.58) by constants  $a_{ij}(\mathbf{0})$ , which is what we shall now assume. Kinetic energy is always positive except in the case that all  $\dot{q}_i$  are equal to 0, and therefore, the quadratic form (7.58) is positive definite.

Motion in a broad class of mechanical systems (so-called *natural systems*) is described by a rather complex system of differential equations—*second-order Lagrange equations*:

$$\frac{d}{dt} \left( \frac{\partial T}{\partial \dot{q}_i} \right) - \frac{\partial T}{\partial q_i} = - \frac{\partial \Pi}{\partial q_i}, \quad i = 1, \dots, n. \quad (7.59)$$

Application of Theorem 7.39 makes it possible to reduce these equations in the given situation to much simpler ones. To this end, let us find a coordinate system in which the quadratic form  $\sum_{i,j=1}^n a_{ij} x_i x_j$  can be brought into the form  $\sum_{i=1}^n x_i^2$ , and the quadratic form  $\sum_{i,j=1}^n b_{ij} x_i x_j$  into the form  $\sum_{i=1}^n \lambda_i x_i^2$ . Then in this case, the form  $\sum_{i,j=1}^n b_{ij} x_i x_j$  is positive definite, which implies that all  $\lambda_i$  are positive. In this system of coordinates (we shall again denote them by  $q_1, \dots, q_n$ ), the system of equations (7.59) is decomposed into the independent equations

$$\frac{d^2 q_i}{dt^2} = -\lambda_i q_i, \quad i = 1, \dots, n, \quad (7.60)$$

which have the solutions  $q_i = c_i \cos \sqrt{\lambda_i} t + d_i \sin \sqrt{\lambda_i} t$ , where  $c_i$  and  $d_i$  are arbitrary constants. This shows that “small oscillations” are periodic in each coordinate  $q_i$ . Since they are bounded, it follows that our equilibrium position  $\mathbf{0}$  is *stable*. If we were to examine the state of equilibrium at a point that was a critical point of



potential energy  $\Pi$  but not a strict minimum, then in the equations (7.60) we would not be able to guarantee that all the  $\lambda_i$  were positive. Then for those  $i$  for which  $\lambda_i < 0$ , we would obtain the solutions  $q_i = c_i \cosh \sqrt{-\lambda_i}t + d_i \sinh \sqrt{-\lambda_i}t$ , which can grow without bound with the growth of  $t$ . Just as for  $\lambda_i = 0$ , we would obtain an unbounded solution  $q_i = c_i + d_i t$ .

Strictly speaking, we have done only the following altogether: we have replaced the given conditions of our problem with conditions close to them, with the result that the problem became much simpler. Such a procedure is usual in the theory of differential equations, where it is proved that solutions to a simplified system of equations are in a certain sense similar to the solutions of the initial system. And moreover, the degree of this deviation can be estimated as a function of the values of the terms that we have ignored. This estimation takes place in a finite interval of time whose length also depends on the value of the ignored terms. This justifies the simplifications that we have made.

A beautiful example, which played an important role historically, is given by lateral oscillations of a string of beads.<sup>4</sup>

Suppose we have a weightless and ideally flexible thread fixed at the ends. On it are securely fastened  $n$  beads with masses  $m_1, \dots, m_n$ , and suppose they divide the thread into segments of lengths  $l_0, l_1, \dots, l_n$ . We shall assume that in its initial state, the thread lies along the  $x$ -axis, and we shall denote by  $y_1, \dots, y_n$  the displacements of the beads along the  $y$ -axis. Then the kinetic energy of this system has the form

$$T = \frac{1}{2} \sum_{i=1}^n m_i \dot{y}_i^2.$$

Assuming the tension of the thread to be constant (as we may because the displacements are small) and equal to  $\sigma$ , we obtain for the potential energy the expression  $\Pi = \sigma \Delta l$ , where  $\Delta l = \sum_{i=0}^n \Delta l_i$  is the change in length of the entire thread, and  $\Delta l_i$  is the change in length of the portion of the thread corresponding to  $l_i$ . Then we know the  $\Delta l_i$  in terms of the  $l_i$ :

$$\Delta l_i = \sqrt{l_i^2 + (y_{i+1} - y_i)^2} - l_i, \quad i = 0, \dots, n,$$

where  $y_0 = y_{n+1} = 0$ . Expanding this expression as a sum in  $y_{i+1} - y_i$ , we obtain quadratic terms  $\sum_{i=0}^n \frac{1}{2l_i} (y_{i+1} - y_i)^2$ , and we may set

$$\Pi = \frac{\sigma}{2} \sum_{i=0}^n \frac{1}{l_i} (y_{i+1} - y_i)^2, \quad y_0 = y_{n+1} = 0.$$

---

<sup>4</sup>This example is taken from Gantmacher and Krein's book *Oscillation Matrices and Kernels and Small Vibrations of Mechanical Systems*, Moscow 1950, English translation, AMS Chelsea Publishing, 2002.

Thus in this case, the problem is reduced to simultaneously expressing two quadratic forms in the variables  $y_1, \dots, y_n$  as sums of squares:

$$T = \frac{1}{2} \sum_{i=0}^n m_i \dot{y}_i^2, \quad \Pi = \frac{\sigma}{2} \sum_{i=0}^n \frac{1}{l_i} (y_{i+1} - y_i)^2, \quad y_0 = y_{n+1} = 0.$$

But if the masses of all the beads are equal and they divide the thread into equal segments, that is,  $m_i = m$  and  $l_i = l/(n+1)$ ,  $i = 1, \dots, n$ , then all the formulas can be written in a more explicit form. In this case, we are speaking about the simultaneous representation as the sum of squares of two forms:

$$T = \frac{m}{2} \sum_{i=1}^n \dot{y}_i^2, \quad \Pi = \frac{\sigma(n+1)}{l} \left( \sum_{i=1}^n y_i^2 - \sum_{i=0}^n y_i y_{i+1} \right), \quad y_0 = y_{n+1} = 0.$$

Therefore, we must use an orthogonal transformation (preserving the form  $\sum_{i=1}^n y_i^2$ ) to express as a sum of squares the form  $\sum_{i=0}^n y_i y_{i+1}$  with matrix

$$A = \frac{1}{2} \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 1 & \ddots & 0 & 0 \\ 0 & 1 & 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & 1 & 0 & 1 \\ 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

It would have been possible to take the standard route: find the eigenvalues  $\lambda_1, \dots, \lambda_n$  as roots of the determinant  $|A - tE|$  and eigenvectors  $\mathbf{y}$  from the system of equations

$$A\mathbf{y} = \lambda\mathbf{y}, \quad (7.61)$$

where  $\lambda = \lambda_i$  and  $\mathbf{y}$  is the column of unknowns  $y_1, \dots, y_n$ . But it is simpler to use equations (7.61) directly. They give a system of  $n$  equations in the unknowns  $y_1, \dots, y_n$ :

$$\begin{aligned} y_2 &= 2\lambda y_1, & y_1 + y_3 &= 2\lambda y_2, & \dots, \\ y_{n-2} + y_n &= 2\lambda y_{n-1}, & y_{n-1} &= 2\lambda y_n, \end{aligned}$$

which can be written in the form

$$y_{k-1} + y_{k+1} = 2\lambda y_k, \quad k = 1, \dots, n, \quad (7.62)$$

where we set  $y_0 = y_{n+1} = 0$ . The system of equations (7.62) is called a *recurrence relation*, whereby each value  $y_{k+1}$  is expressed in terms of the two preceding values:  $y_k$  and  $y_{k-1}$ . Thus if we know two adjacent values, then we can use relationship

(7.62) to construct all the  $y_k$ . The condition  $y_0 = y_{n+1} = 0$  is called a *boundary condition*.

Let us note that for  $\lambda = \pm 1$ , the equation (7.62) with boundary condition  $y_0 = y_{n+1} = 0$  has only the null solution:  $y_0 = \dots = y_{n+1} = 0$ . Indeed, for  $\lambda = 1$ , we obtain

$$y_2 = 2y_1, \quad y_3 = 3y_1, \quad \dots, \quad y_n = ny_1, \quad y_{n+1} = (n+1)y_1,$$

from which by  $y_{n+1} = 0$  it follows that  $y_1 = 0$ , and all  $y_k$  are equal to 0. Similarly, for  $\lambda = -1$ , we obtain

$$\begin{aligned} y_2 &= -2y_1, & y_3 &= 3y_1, & y_4 &= -4y_1, & \dots, \\ y_n &= (-1)^{n-1}ny_1, & y_{n+1} &= (-1)^n(n+1)y_1, \end{aligned}$$

from which by  $y_{n+1} = 0$  it follows as well that  $y_1 = 0$ , and again all the  $y_k$  are equal to zero. Thus for  $\lambda = \pm 1$ , the system of equations (7.61) has as its only solution the vector  $\mathbf{y} = \mathbf{0}$ , which by definition, cannot be an eigenvector. In other words, this implies that the numbers  $\pm 1$  are not eigenvalues of the matrix  $A$ .

There is a lovely formula for solving equation (7.62) with boundary condition  $y_0 = y_{n+1} = 0$ . Let us denote by  $\alpha$  and  $\beta$  the roots of the quadratic equation  $z^2 - 2\lambda z + 1 = 0$ . By the above reasoning,  $\lambda \neq \pm 1$ , and therefore, the numbers  $\alpha$  and  $\beta$  are distinct and cannot equal  $\pm 1$ . Direct substitution shows that then for arbitrary  $A$  and  $B$ , the sequence  $y_k = A\alpha^k + B\beta^k$  satisfies the relationship (7.62). The coefficients  $A$  and  $B$  taken to satisfy  $y_0 = 0$ ,  $y_1$  are given. The following  $y_k$ , as we have seen, are determined by the relationship (7.62), and this implies that again they are given by our formula. The conditions  $y_0 = 0$ ,  $y_1$  fixed give  $B = -A$  and  $A(\alpha - \beta) = y_1$ , whence  $A = y_1/(\alpha - \beta)$ . Thus we obtain the expression

$$y_k = \frac{y_1}{\alpha - \beta}(\alpha^k - \beta^k). \quad (7.63)$$

We now use the condition  $y_{n+1} = 0$ , which gives  $\alpha^{n+1} = \beta^{n+1}$ . Moreover, since  $\alpha$  and  $\beta$  are roots of the polynomial  $z^2 - 2\lambda z + 1$ , we have  $\alpha\beta = 1$ , whence  $\beta = \alpha^{-1}$ , which implies that  $\alpha^{2(n+1)} = 1$ . From this (taking into account that  $\alpha \neq \pm 1$ ), we obtain

$$\alpha = \cos\left(\frac{\pi j}{n+1}\right) + i \sin\left(\frac{\pi j}{n+1}\right),$$

where  $i$  is the imaginary unit, and the number  $j$  assumes the values  $1, \dots, n$ . Again using the equation  $z^2 - 2\lambda z + 1 = 0$ , whose roots are  $\alpha$  and  $\beta$ , we obtain  $n$  distinct values for  $\lambda$ :

$$\lambda_j = \cos\left(\frac{\pi j}{n+1}\right), \quad j = 1, \dots, n,$$

since  $j = n+2, \dots, 2n+1$  give the same values  $\lambda_j$ . These are precisely the eigenvalues of the matrix  $A$ . For the eigenvector  $\mathbf{y}_j$  of the associated eigenvalue  $\lambda_j$ , we

obtain by formula (7.63) its coordinates  $y_{1j}, \dots, y_{nj}$  in the form

$$y_{kj} = \sin\left(\frac{\pi kj}{n+1}\right), \quad k = 1, \dots, n.$$

These formulas were derived by d'Alembert and Daniel Bernoulli. Passing to the limit as  $n \rightarrow \infty$ , Lagrange derived from these the law of vibrations of a uniform string.

*Example 7.45* Let us consider in an  $n$ -dimensional real Euclidean space  $L$  the subset  $X$  given by the equation

$$F(x_1, \dots, x_n) = 0 \tag{7.64}$$

in some coordinate system. Such a subset  $X$  is called a *hypersurface* and consists of all vectors  $\mathbf{x} = (x_1, \dots, x_n)$  of the Euclidean space  $L$  whose coordinates satisfy the equation<sup>5</sup> (7.64). Using the change-of-coordinates formula (3.36), we see that the property of the subset  $X \subset L$  being a hypersurface does not depend on the choice of coordinates, that is, on the choice of the basis of  $L$ . Then if we assume that the beginning of every vector is located at a single fixed point, then every vector  $\mathbf{x} = (x_1, \dots, x_n)$  can be identified with its endpoint, a point of the given space. In order to conform to more customary terminology, as we continue with this example, we shall call the vectors  $\mathbf{x}$  of which the hypersurface  $X$  consists its *points*.

We shall assume that  $F(\mathbf{0}) = 0$  and that the function  $F(x_1, \dots, x_n)$  is differentiable in each of its arguments as many times as necessary. It is easily verified that this condition also does not depend on the choice of basis. Let us assume in addition that  $\mathbf{0}$  is not a critical point of the hypersurface  $X$ , that is, that not all partial derivatives  $\partial F(\mathbf{0})/\partial x_i$  are equal to zero. In other words, if we introduce the vector  $\text{grad } F = (\partial F/\partial x_1, \dots, \partial F/\partial x_n)$ , called the *gradient* of the function  $F$ , then this implies that  $\text{grad } F(\mathbf{0}) \neq \mathbf{0}$ .

We shall be interested in *local* properties of the hypersurface  $X$ , that is, properties associated with points close to  $\mathbf{0}$ . With the assumptions that we have made, the *implicit function theorem*, known from analysis, shows that near  $\mathbf{0}$ , the coordinates  $x_1, \dots, x_n$  of each point of the hypersurface  $X$  can be represented as a function of  $n-1$  arguments  $u_1, \dots, u_{n-1}$ , and furthermore, for each point, the values  $u_1, \dots, u_{n-1}$  are uniquely determined. It is possible to choose as  $u_1, \dots, u_{n-1}$  some  $n-1$  of the coordinates  $x_1, \dots, x_n$ , after determining the remaining coordinate  $x_k$  from equation (7.64), for which must be satisfied only the condition  $\frac{\partial F}{\partial x_k}(\mathbf{0}) \neq 0$  for the given  $k$ , which holds because of the assumption  $\text{grad } F(\mathbf{0}) \neq \mathbf{0}$ . The functions that determine the dependence of the coordinates  $x_1, \dots, x_n$  of a point of the hyperplane  $X$  on the arguments  $u_1, \dots, u_{n-1}$  are differentiable at all arguments as many times as the original function  $F(x_1, \dots, x_n)$ .

---

<sup>5</sup>The more customary point of view, when the hypersurface (for example, a curve or surface) consists of *points*, requires the consideration of an  $n$ -dimensional space consisting of *points* (otherwise *affine* space), which will be introduced in the following chapter.

The *hyperplane* defined by the equation

$$\sum_{i=1}^n \frac{\partial F}{\partial x_i}(\mathbf{0})x_i = 0$$

is called the *tangent space* or *tangent hyperplane* to the hypersurface  $X$  at the point  $\mathbf{0}$  and is denoted by  $T_0X$ . In the case that the basis of the Euclidean space  $L$  is orthonormal, this equation can also be written in the form  $(\text{grad } F(\mathbf{0}), \mathbf{x}) = 0$ . As a subspace of the Euclidean space  $L$ , the tangent space  $T_0X$  is also a Euclidean space.

The set of vectors depending on the parameter  $t$  taking values on some interval of the real line, that is,  $\mathbf{x}(t) = (x_1(t), \dots, x_n(t))$ , is called a *smooth curve* if all functions  $x_i(t)$  are differentiable a sufficient number of times and if for every value of the parameter  $t$ , not all the derivatives  $dx_i/dt$  are equal to zero. In analogy to what was said above about hypersurfaces, we may visualize the curve as consisting of *points*  $A(t)$ , where each  $A(t)$  is the endpoint of some vector  $\mathbf{x}(t)$ , while all the vectors  $\mathbf{x}(t)$  begin at a certain fixed point  $O$ . In what follows, we shall refer to the vectors  $\mathbf{x}$  that constitute the curve as its *points*.

We say that a curve  $\gamma$  *passes through* the point  $\mathbf{x}_0$  if  $\mathbf{x}(t_0) = \mathbf{x}_0$  for some value of the parameter  $t_0$ . It is clear that here we may always assume that  $t_0 = 0$ . Indeed, let us consider a different curve  $\tilde{\mathbf{x}}(t) = (\tilde{x}_1(t), \dots, \tilde{x}_n(t))$ , where the functions  $\tilde{x}_i(t)$  are equal to  $x_i(t + t_0)$ . This can also be written in the form  $\tilde{\mathbf{x}}(\tau) = \mathbf{x}(t)$ , where we have introduced a new parameter  $\tau$  related to the old one by  $\tau = t - t_0$ .

Generally speaking, for a curve we may make an arbitrary *change of parameter* by the formula  $t = \psi(\tau)$ , where the function  $\psi$  defines a continuously differentiable bijective mapping of one interval to another. Under such a change, a curve, considered as a set of points (or vectors), will remain the same. From this it follows that one and the same curve can be written in a variety of ways using various parameters.<sup>6</sup>

We now introduce the vector  $\frac{d\mathbf{x}}{dt} = (\frac{dx_1}{dt}, \dots, \frac{dx_n}{dt})$ . Suppose the curve  $\gamma$  passes through the point  $\mathbf{0}$  for  $t = 0$ . Then the vector  $\mathbf{p} = \frac{d\mathbf{x}}{dt}(\mathbf{0})$  is called a *tangent vector* to the curve  $\gamma$  at the point  $\mathbf{0}$ . It depends, of course, on the choice of parameter  $t$  defining the curve. Under a change of parameter  $t = \psi(\tau)$ , we have

$$\frac{d\mathbf{x}}{d\tau} = \frac{d\mathbf{x}}{dt} \cdot \frac{dt}{d\tau} = \frac{d\mathbf{x}}{dt} \cdot \psi'(\tau), \quad (7.65)$$

and the tangent vector  $\mathbf{p}$  is multiplied by a constant equal to the value of the derivative  $\psi'(0)$ . Using this fact, it is possible to arrange things so that  $|\frac{d\mathbf{x}}{dt}(t)| = 1$  for all  $t$  close to 0. Such a parameter is said to be *natural*. The condition that the curve  $\mathbf{x}(t)$  belong to the hyperplane (7.64) gives the equality  $F(\mathbf{x}(t)) = 0$ , which is satisfied for all  $t$ . Differentiating this relationship with respect to  $t$ , we obtain that the vector  $\mathbf{p}$  lies in the space  $T_0X$ . And conversely, an arbitrary vector contained in  $T_0X$  can

<sup>6</sup>For example, the circle of radius 1 with center at the origin with Cartesian coordinates  $x, y$  can be defined not only by the formula  $x = \cos t, y = \sin t$ , but also by the formula  $x = \cos \tau, y = -\sin \tau$  (with the replacement  $t = -\tau$ ), or by the formula  $x = \sin \tau, y = \cos \tau$  (replacement  $t = \frac{\pi}{2} - \tau$ ).

be represented in the form  $\frac{dx}{dt}(0)$  for some curve  $\mathbf{x}(t)$ . This curve, of course, is not uniquely determined. Curves whose tangent vectors  $\mathbf{p}$  are proportional are said to be *tangent* at the point  $\mathbf{0}$ .

Let us denote by  $\mathbf{n}$  a unit vector orthogonal to the tangent space  $T_{\mathbf{0}}X$ . There are two such vectors,  $\mathbf{n}$  and  $-\mathbf{n}$ , and we shall choose one of them. For example, we may set

$$\mathbf{n} = \frac{\text{grad } F}{|\text{grad } F|}(\mathbf{0}). \quad (7.66)$$

We define the vector  $\frac{d^2\mathbf{x}}{dt^2}$  as  $\frac{d}{dt}(\frac{d\mathbf{x}}{dt})$  and set

$$Q = \left( \frac{d^2\mathbf{x}}{dt^2}(0), \mathbf{n} \right). \quad (7.67)$$

**Proposition 7.46** *The value  $Q$  depends only on the vector  $\mathbf{p}$ ; namely, it is a quadratic form in its coordinates.*

*Proof* It suffices to verify this assertion by substituting in (7.67) for the vector  $\mathbf{n}$ , any vector proportional to it, for example,  $\text{grad } F(\mathbf{0})$ . Since by assumption, the curve  $\mathbf{x}(t)$  is contained in the hyperplane (7.64), it follows that  $F(x_1(t), \dots, x_n(t)) = 0$ . Differentiating this equality twice with respect to  $t$ , we obtain

$$\sum_{i=1}^n \frac{\partial F}{\partial x_i} \frac{dx_i}{dt} = 0, \quad \sum_{i,j=1}^n \frac{\partial^2 F}{\partial x_i \partial x_j} \frac{dx_i}{dt} \frac{dx_j}{dt} + \sum_{i=1}^n \frac{\partial F}{\partial x_i} \frac{d^2x_i}{dt^2} = 0.$$

Setting here  $t = 0$ , we see that

$$\left( \frac{d^2\mathbf{x}}{dt^2}(0), \text{grad } F(\mathbf{0}) \right) = - \sum_{i,j=1}^n \frac{\partial^2 F}{\partial x_i \partial x_j}(\mathbf{0}) p_i p_j,$$

where  $\mathbf{p} = (p_1, \dots, p_n)$ . This proves the assertion.  $\square$

The form  $Q(\mathbf{p})$  is called the *second quadratic form* of the hypersurface. The form  $(\mathbf{p}^2)$  is called the *first quadratic form* when  $T_{\mathbf{0}}X$  is taken as a subspace of a Euclidean space  $L$ . We observe that the second quadratic form requires the selection of one of two unit vectors ( $\mathbf{n}$  or  $-\mathbf{n}$ ) orthogonal to  $T_{\mathbf{0}}X$ . This is frequently interpreted as the selection of *one side* of the hypersurface in a neighborhood of the point  $\mathbf{0}$ .

The first and second quadratic forms give us the possibility to obtain an expression for the curvature of certain curves  $\mathbf{x}(t)$  lying in the hypersurface  $X$ . Let us suppose that a curve is the intersection of a plane  $L'$  containing the point  $\mathbf{0}$  and the hypersurface  $X$  (even if only in an arbitrarily small neighborhood of the point  $\mathbf{0}$ ). Such a curve is called a *plane section* of the hypersurface. If we define the curve  $\mathbf{x}(t)$  in such a way that  $t$  is a natural parameter, then its *curvature* at the point  $\mathbf{0}$  is

the number

$$k = \left| \frac{d^2 \mathbf{x}}{dt^2}(0) \right|.$$

We assume that  $k \neq 0$  and set

$$\mathbf{m} = \frac{1}{k} \cdot \frac{d^2 \mathbf{x}}{dt^2}(0).$$

The vector  $\mathbf{m}$  has length 1 by definition. It is said to be *normal* to the curve  $\mathbf{x}(t)$  at the point  $\mathbf{0}$ . If the curve  $\mathbf{x}(t)$  is a plane section of the hypersurface, then  $\mathbf{x}(t)$  lies in the plane  $L'$  (for all sufficiently small  $t$ ), and consequently, the vector

$$\frac{d\mathbf{x}}{dt} = \lim_{h \rightarrow 0} \frac{\mathbf{x}(t+h) - \mathbf{x}(t)}{h}$$

also lies in the plane  $L'$ . Therefore, this holds as well for the vector  $d^2 \mathbf{x}/dt^2$ , which implies that it holds as well for the normal  $\mathbf{m}$ . If the curve  $\gamma$  is defined in terms of the natural parameter  $t$ , then

$$\left| \frac{d\mathbf{x}}{dt} \right|^2 = \left( \frac{d\mathbf{x}}{dt}, \frac{d\mathbf{x}}{dt} \right) = 1.$$

Differentiating this equality with respect to  $t$ , we obtain that the vectors  $d^2 \mathbf{x}/dt^2$  and  $d\mathbf{x}/dt$  are orthogonal. Hence the normal  $\mathbf{m}$  to the curve  $\gamma$  is orthogonal to an arbitrary tangent vector (for arbitrary definition of the curve  $\gamma$  in the form  $\mathbf{x}(t)$  with natural parameter  $t$ ), and the vector  $\mathbf{m}$  is defined uniquely up to sign. It is obvious that  $L' = \langle \mathbf{m}, \mathbf{p} \rangle$ , where  $\mathbf{p}$  is an arbitrary tangent vector.

By definition (7.67) of the second quadratic form  $Q$  and taking into account the equality  $|\mathbf{m}| = |\mathbf{n}| = 1$ , we obtain the expression

$$Q(\mathbf{p}) = (k\mathbf{m}, \mathbf{n}) = k(\mathbf{m}, \mathbf{n}) = k \cos \varphi, \quad (7.68)$$

where  $\varphi$  is the angle between the vectors  $\mathbf{m}$  and  $\mathbf{n}$ . The expression  $k \cos \varphi$  is denoted by  $\tilde{k}$  and is called the *normal curvature* of the hypersurface  $X$  in the direction  $\mathbf{p}$ . We recall that here  $\mathbf{n}$  denotes the chosen unit vector orthogonal to the tangent space  $T_0 X$ , and  $\mathbf{m}$  is the normal to the curve to which the vector  $\mathbf{p}$  is tangent. An analogous formula for an arbitrary parametric definition of the curve  $\mathbf{x}(t)$  (where  $t$  is not necessarily a natural parameter) also uses the first quadratic form. Namely, if  $\tau$  is another parameter, while  $t$  is a natural parameter, then by formula (7.65), now instead of the vector  $\mathbf{p}$ , we obtain  $\mathbf{p}' = \mathbf{p}\psi'(0)$ . Since  $Q$  is a quadratic form, it follows that  $Q(\mathbf{p}\psi'(0)) = \psi'(0)^2 Q(\mathbf{p})$ , and instead of formula (7.68), we now obtain

$$\frac{Q(\mathbf{p})}{(\mathbf{p}^2)} = k \cos \varphi. \quad (7.69)$$

Here the first quadratic form  $(\mathbf{p}^2)$  is already involved as well as the second quadratic form  $Q(\mathbf{p})$ , but now (7.69), in contrast to (7.68), holds for an arbitrary choice of parameter  $t$  on the curve  $\gamma$ .

The point of the term *normal curvature* given above is the following. The section of the hypersurface  $X$  by the plane  $L'$  is said to be *normal* if  $\mathbf{n} \in L'$ . The vector  $\mathbf{n}$  defined by formula (7.66) is orthogonal to the tangent plane  $T_0X$ . But in the plane  $L'$  there is also the vector  $\mathbf{p}$  tangent to the curve  $\gamma$ , and the normal vector  $\mathbf{m}$  orthogonal to it. Thus in the case of a normal section  $\mathbf{n} = \pm\mathbf{m}$ , this means that in formula (7.68), the angle  $\varphi$  is equal to 0 or  $\pi$ . Conversely, from the equality  $|\cos \varphi| = 1$ , it follows that  $\mathbf{n} \in L'$ . Thus in the case of a normal section, the normal curvature  $\tilde{k}$  differs from  $k$  only by the factor  $\pm 1$  and is defined by the relationship

$$\tilde{k} = \frac{Q(\mathbf{p})}{|\mathbf{p}|^2}.$$

Since  $L' = \langle \mathbf{m}, \mathbf{p} \rangle$ , it follows that all normal sections correspond to straight lines in the plane  $L'$ . For each line, there exists a unique normal section containing this line. In other words, we “rotate” the plane  $L'$  about the vector  $\mathbf{m}$ , considering all obtained planes  $\langle \mathbf{m}, \mathbf{p} \rangle$ , where  $\mathbf{p}$  is a vector in the tangent hyperplane  $T_0X$ . Thus all normal sections of the hypersurface  $X$  are obtained.

We shall now employ Theorem 7.38. In our case, it gives an orthonormal basis  $\mathbf{e}_1, \dots, \mathbf{e}_{n-1}$  in the tangent hyperplane  $T_0X$  (viewed as a subspace of the Euclidean space  $L$ ) in which the quadratic form  $Q(\mathbf{p})$  is brought into canonical form. In other words, for the vector  $\mathbf{p} = u_1\mathbf{e}_1 + \dots + u_{n-1}\mathbf{e}_{n-1}$ , the second quadratic form takes the form  $Q(\mathbf{p}) = \lambda_1 u_1^2 + \dots + \lambda_{n-1} u_{n-1}^2$ . Since the basis  $\mathbf{e}_1, \dots, \mathbf{e}_{n-1}$  is orthonormal, we have in this case

$$\frac{u_i}{|p_i|} = \frac{(\mathbf{p}, \mathbf{e}_i)}{|p_i|} = \cos \varphi_i, \quad (7.70)$$

where  $\varphi_i$  is the angle between the vectors  $\mathbf{p}$  and  $\mathbf{e}_i$ . From this we obtain for the normal curvature  $\tilde{k}$  of the normal section  $\gamma$ , the formula

$$\tilde{k} = \frac{Q(\mathbf{p})}{|\mathbf{p}|^2} = \sum_{i=1}^{n-1} \lambda_i \left( \frac{u_i}{|p_i|} \right)^2 = \sum_{i=1}^{n-1} \lambda_i \cos^2 \varphi_i, \quad (7.71)$$

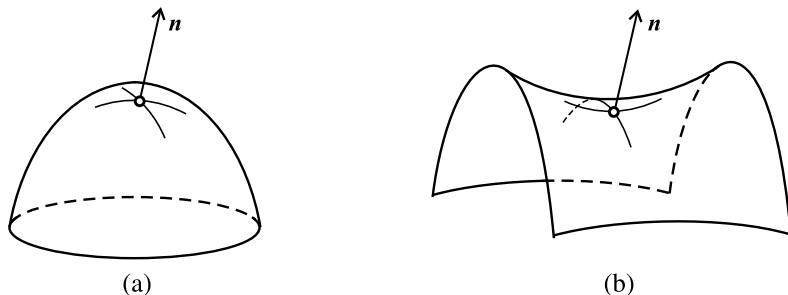
where  $\mathbf{p}$  is an arbitrary tangent vector to the curve  $\gamma$  at the point  $\mathbf{0}$ . Relationships (7.70) and (7.71) are called *Euler's formula*. The numbers  $\lambda_i$  are called *principal curvatures* of the hypersurface  $X$  at the point  $\mathbf{0}$ .

In the case  $n = 3$ , the hypersurface (7.64) is an ordinary surface and has two principal curvatures  $\lambda_1$  and  $\lambda_2$ . Taking into account the fact that  $\cos^2 \varphi_1 + \cos^2 \varphi_2 = 1$ , Euler's formula takes the form

$$\tilde{k} = \lambda_1 \cos^2 \varphi_1 + \lambda_2 \cos^2 \varphi_2 = (\lambda_1 - \lambda_2) \cos^2 \varphi_1 + \lambda_2. \quad (7.72)$$

Suppose  $\lambda_1 \geq \lambda_2$ . Then from (7.72), it is clear that the normal curvature  $\tilde{k}$  assumes a maximum (equal to  $\lambda_1$ ) for  $\cos^2 \varphi_1 = 1$  and a minimum (equal to  $\lambda_2$ ) for





**Fig. 7.10** Elliptic (a) and hyperbolic (b) points

$\cos^2 \varphi_1 = 0$ . This assertion is called the *extremal property* of the principal curvatures of the surface. If  $\lambda_1$  and  $\lambda_2$  have the same sign ( $\lambda_1 \lambda_2 > 0$ ), then as can be seen from (7.72), an arbitrary normal section of a surface at a given point  $\mathbf{0}$  has its curvature of the same sign, and therefore, all normal sections have convexity in the same direction, and near the point  $\mathbf{0}$ , the surface lies *on one side* of its tangent plane; see Fig. 7.10(a). Such points are called *elliptic*. If  $\lambda_1$  and  $\lambda_2$  have different signs ( $\lambda_1 \lambda_2 < 0$ ), then as can be seen from formula (7.72), there exist normal sections with opposite directions of convexity, and at points near  $\mathbf{0}$ , the surface is located *on different sides* of its tangent plane; see Fig. 7.10(b). Such points are called *hyperbolic*.<sup>7</sup>

From all this discussion, it is evident that the product of principal curvatures  $\kappa = \lambda_1 \lambda_2$  characterizes some important properties of a surface (called “internal geometric properties” of the surface). This product is called the *Gaussian* or *total curvature* of the surface.

## 7.7 Pseudo-Euclidean Spaces

Many of the theorems proved in the previous sections of this chapter remain valid if in the definition of Euclidean space we forgo the requirement of positive definiteness of the quadratic form ( $\mathbf{x}^2$ ) or replace it with something weaker. Without this condition, the inner product  $(\mathbf{x}, \mathbf{y})$  does not differ at all from an arbitrary symmetric bilinear form. As Theorem 6.6 shows, it is uniquely defined by the quadratic form ( $\mathbf{x}^2$ ).

We thus obtain a theory that fully coincides with the theory of quadratic forms that we presented in Chap. 6. The fundamental theorem (on bringing a quadratic form into canonical form) consists in the existence of an orthonormal basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , that is, a basis for which  $(\mathbf{e}_i, \mathbf{e}_j) = 0$  for all  $i \neq j$ . Then for the vector  $x_1 \mathbf{e}_1 + \dots + x_n \mathbf{e}_n$ , the quadratic form ( $\mathbf{x}^2$ ) is equal to  $\lambda_1 x_1^2 + \dots + \lambda_n x_n^2$ .

<sup>7</sup>Examples of surfaces consisting entirely of elliptic points are ellipsoids, hyperboloids of two sheets, and elliptic paraboloids, while surfaces consisting entirely of hyperbolic points include hyperboloids of one sheet and hyperbolic paraboloids.

Moreover, this is true for vector spaces and bilinear forms over an arbitrary field  $\mathbb{K}$  of characteristic different from 2. The concept of an isomorphism of spaces makes sense also in this case; as previously, it is necessary to require that the scalar product  $(\mathbf{x}, \mathbf{y})$  be preserved.

The theory of such spaces (defined up to isomorphism) with a bilinear or quadratic form is of great interest (for example, in the case  $\mathbb{K} = \mathbb{Q}$ , the field of rational numbers). But here we are interested in real spaces. In this case, formula (6.28) and Theorem 6.17 (law of inertia) show that up to isomorphism, a space is uniquely defined by its rank and the index of inertia of the associated quadratic form.

We shall further restrict attention to an examination of real vector spaces with a nonsingular symmetric bilinear form  $(\mathbf{x}, \mathbf{y})$ . Let us recall that the nonsingularity of a bilinear form implies that its rank (that is, the rank of its matrix in an arbitrary basis of the space) is equal to  $\dim L$ . In other words, this means that its radical is equal to  $(\mathbf{0})$ ; that is, if the vector  $\mathbf{x}$  is such that  $(\mathbf{x}, \mathbf{y}) = 0$  for all vectors  $\mathbf{y} \in L$ , then  $\mathbf{x} = \mathbf{0}$  (see Sect. 6.2). For a Euclidean space, this condition follows automatically from property (4) of the definition (it suffices to set there  $\mathbf{y} = \mathbf{x}$ ).

Formula (6.28) shows that with these conditions, there exists a basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the space  $L$  for which

$$(\mathbf{e}_i, \mathbf{e}_j) = 0 \quad \text{for } i \neq j, \quad (\mathbf{e}_i^2) = \pm 1.$$

Such a basis is called, as it was previously, *orthonormal*. In it, the form  $(\mathbf{x}^2)$  can be written in the form

$$(\mathbf{x}^2) = x_1^2 + \dots + x_s^2 - x_{s+1}^2 - \dots - x_n^2,$$

and the number  $s$  is called the *index of inertia* of both the quadratic form  $(\mathbf{x}^2)$  and the pseudo-Euclidean space  $L$ .

A new difficulty appears that was not present for Euclidean spaces if the quadratic form  $(\mathbf{x}^2)$  is neither positive nor negative definite, that is, if its index of inertia  $s$  is positive but less than  $n$ . In this case, the restriction of the bilinear form  $(\mathbf{x}, \mathbf{y})$  to the subspace  $L' \subset L$  can turn out to be singular, even if the original bilinear form  $(\mathbf{x}, \mathbf{y})$  in  $L$  was nonsingular. For example, it is clear that in  $L$ , there exists a vector  $\mathbf{x} \neq \mathbf{0}$  for which  $(\mathbf{x}^2) = 0$ , and then the restriction of  $(\mathbf{x}, \mathbf{y})$  to a one-dimensional subspace  $\langle \mathbf{x} \rangle$  is singular (identically equal to zero).

Thus let us consider a vector space  $L$  with a nonsingular symmetric bilinear form  $(\mathbf{x}, \mathbf{y})$  defined on it. In this case, we shall use many concepts and much of the notation used for Euclidean spaces earlier. Hence, vectors  $\mathbf{x}$  and  $\mathbf{y}$  are called *orthogonal* if  $(\mathbf{x}, \mathbf{y}) = 0$ . Subspaces  $L_1$  and  $L_2$  are called *orthogonal* if  $(\mathbf{x}, \mathbf{y}) = 0$  for all vectors  $\mathbf{x} \in L_1$  and  $\mathbf{y} \in L_2$ , and we express this by writing  $L_1 \perp L_2$ . The orthogonal complement of the subspace  $L' \subset L$  with respect to the bilinear form  $(\mathbf{x}, \mathbf{y})$  is denoted by  $(L')^\perp$ . However, there is an important difference from the case of Euclidean spaces, in connection with which it will be useful to give the following definition.

**Definition 7.47** A subspace  $L' \subset L$  is said to be *nondegenerate* if the bilinear form obtained by restricting the form  $(\mathbf{x}, \mathbf{y})$  to  $L'$  is nonsingular. In the contrary case,  $L'$  is said to be *degenerate*.

By Theorem 6.9, in the case of a nondegenerate subspace  $L'$  we have the orthogonal decomposition

$$L = L' \oplus (L')^\perp. \quad (7.73)$$

In the case of a Euclidean space, as we have seen, every subspace  $L'$  is nondegenerate, and the decomposition (7.73) holds without any additional conditions. As the following example will show, in a pseudo-Euclidean space, the condition of nondegeneracy of a subspace  $L'$  for the decomposition (7.73) is in fact essential.

*Example 7.48* Let us consider a three-dimensional space  $L$  with a symmetric bilinear form defined in some chosen basis by the formula

$$(\mathbf{x}, \mathbf{y}) = x_1 y_1 + x_2 y_2 - x_3 y_3,$$

where the  $x_i$  are the coordinates of the vector  $\mathbf{x}$ , and the  $y_i$  are the coordinates of the vector  $\mathbf{y}$ . Let  $L' = \langle \mathbf{e} \rangle$ , where the vector  $\mathbf{e}$  has coordinates  $(0, 1, 1)$ . Then as is easily verified,  $(\mathbf{e}, \mathbf{e}) = 0$ , and therefore, the restriction of the form  $(\mathbf{x}, \mathbf{y})$  to  $L'$  is identically equal to zero. This implies that the subspace  $L'$  is degenerate. Its orthogonal complement  $(L')^\perp$  is two-dimensional and consists of all vectors  $\mathbf{z} \in L$  with coordinates  $(z_1, z_2, z_3)$  for which  $z_2 = z_3$ . Consequently,  $L' \subset (L')^\perp$ , and the intersection  $L' \cap (L')^\perp = L'$  contains nonnull vectors. This implies that the sum  $L' + (L')^\perp$  is not a direct sum. Furthermore, it is obvious that  $L' + (L')^\perp \neq L$ .

It follows from the nonsingularity of a bilinear form  $(\mathbf{x}, \mathbf{y})$  that the determinant of its matrix (in an arbitrary basis) is different from zero. If this matrix is written in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , then its determinant is equal to

$$\begin{vmatrix} (\mathbf{e}_1, \mathbf{e}_1) & (\mathbf{e}_1, \mathbf{e}_2) & \cdots & (\mathbf{e}_1, \mathbf{e}_n) \\ (\mathbf{e}_2, \mathbf{e}_1) & (\mathbf{e}_2, \mathbf{e}_2) & \cdots & (\mathbf{e}_2, \mathbf{e}_n) \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{e}_n, \mathbf{e}_1) & (\mathbf{e}_n, \mathbf{e}_2) & \cdots & (\mathbf{e}_n, \mathbf{e}_n) \end{vmatrix}, \quad (7.74)$$

and just as in the case of a Euclidean space, we shall call this its *Gram determinant* of the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ . Of course, this determinant depends on the choice of basis, but its *sign* does not depend on the basis. Indeed, if  $A$  and  $A'$  are matrices of our bilinear form in two different bases, then they are related by the equality  $A' = C^* A C$ , where  $C$  is a nonsingular transition matrix, from which it follows that  $|A'| = |A| \cdot |C|^2$ . Thus the sign of the Gram determinant is the same for all bases.

As noted above, for a nondegenerate subspace  $L' \subset L$ , we have the decomposition (7.73), which yields the equality

$$\dim L = \dim L' + \dim (L')^\perp. \quad (7.75)$$

But equality (7.75) holds as well for every subspace  $L' \subset L$ , although as we saw in Example 7.48, the decomposition (7.73) may already not hold in the general case.

Indeed, by Theorem 6.3, we can write an arbitrary bilinear form  $(\mathbf{x}, \mathbf{y})$  in the space  $L$  in the form  $(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathcal{A}(\mathbf{y}))$ , where  $\mathcal{A} : L \rightarrow L^*$  is some linear transformation. From the nonsingularity of the bilinear form  $(\mathbf{x}, \mathbf{y})$  follows the nonsingularity of the transformation  $\mathcal{A}$ . In other words, the transformation  $\mathcal{A}$  is an isomorphism, that is, its kernel is equal to  $(\mathbf{0})$ , and in particular, for an arbitrary subspace  $L' \subset L$ , we have the equality  $\dim \mathcal{A}(L') = \dim L'$ . On the other hand, we can write the orthogonal complement  $(L')^\perp$  in the form  $(\mathcal{A}(L'))^a$ , using the notion of the annihilator introduced in Sect. 3.7. On the basis of what we have said above and formula (3.54) for the annihilator, we have the relationship

$$\dim(\mathcal{A}(L'))^a = \dim L - \dim \mathcal{A}(L') = \dim L - \dim L',$$

that is,  $\dim(L')^\perp = \dim L - \dim L'$ . We note that this argument holds for vector spaces  $L$  defined not only over the real numbers, but over any field.

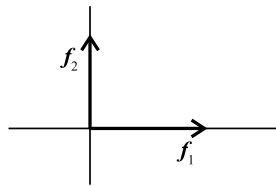
The spaces that we have examined are defined (up to isomorphism) by the index of inertia  $s$ , which can take values from 0 to  $n$ . By what we have said above, the sign of the Gram determinant of an arbitrary basis is equal to  $(-1)^{n-s}$ . It is obvious that if we replace the inner product  $(\mathbf{x}, \mathbf{y})$  in the space  $L$  by  $-(\mathbf{x}, \mathbf{y})$ , we shall preserve all of its essential properties, but the index of inertia  $s$  will be replaced by  $n - s$ , whence in what follows, we shall assume that  $n/2 \leq s \leq n$ . The case  $s = n$  corresponds to a Euclidean space. There exists, however, a phenomenon whose explanation is at present not completely clear; the most interesting questions in mathematics and physics were until now connected with two types of spaces: those in which the index of inertia  $s$  is equal to  $n$  and those for which  $s = n - 1$ . The theory of Euclidean spaces ( $s = n$ ) has been up till now the topic of this chapter. In the remaining part, we shall consider the other case:  $s = n - 1$ . In the sequel, we shall call such spaces *pseudo-Euclidean spaces* (although sometimes, this term is used when  $(\mathbf{x}, \mathbf{y})$  is an arbitrary nonsingular symmetric bilinear form neither positive nor negative definite, that is, with index of inertia  $s \neq 0, n$ ).

Thus a pseudo-Euclidean space of dimension  $n$  is a vector space  $L$  equipped with a symmetric bilinear form  $(\mathbf{x}, \mathbf{y})$  such that in some basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , the quadratic form  $(\mathbf{x}^2)$  takes the form

$$x_1^2 + \dots + x_{n-1}^2 - x_n^2. \quad (7.76)$$

As in the case of a Euclidean space, we shall, as we did previously, call such bases *orthonormal*.

The best-known application of pseudo-Euclidean spaces is related to the *special theory of relativity*. According to an idea put forward by Minkowski, in this theory, one considers a four-dimensional space whose vectors are called *space-time events* (we mentioned this earlier, on p. 86). They have coordinates  $(x, y, z, t)$ , and the space is equipped with a quadratic form  $x^2 + y^2 + z^2 - t^2$  (here the speed of light is assumed to be 1). The pseudo-Euclidean space thus obtained is called *Minkowski space*. By analogy with the physical sense of these concepts in Minkowski space, in an arbitrary pseudo-Euclidean space, a vector  $\mathbf{x}$  is said to be *spacelike* if  $(\mathbf{x}^2) > 0$ ,

**Fig. 7.11** A pseudo-Euclidean plane

while such a vector is said to be *timelike* if  $(x^2) < 0$ , and *lightlike*, or *isotropic*, if  $(x^2) = 0$ .<sup>8</sup>

**Example 7.49** Let us consider the simplest case of a pseudo-Euclidean space  $L$  with  $\dim L = 2$  and index of inertia  $s = 1$ . By the general theory, in this space there exists an orthonormal basis, in this case the basis  $e_1, e_2$ , for which

$$(e_1^2) = 1, \quad (e_2^2) = -1, \quad (e_1, e_2) = 0, \quad (7.77)$$

and the scalar square of the vector  $x = x_1 e_1 + x_2 e_2$  is equal to  $(x^2) = x_1^2 - x_2^2$ . However, it is easier to write the formulas connected with the space  $L$  in the basis consisting of lightlike vectors  $f_1, f_2$ , after setting

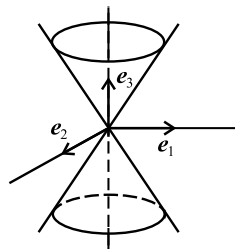
$$f_1 = \frac{e_1 + e_2}{2}, \quad f_2 = \frac{e_1 - e_2}{2}. \quad (7.78)$$

Then  $(f_1^2) = (f_2^2) = 0$ ,  $(f_1, f_2) = \frac{1}{2}$ , and the scalar square of the vector  $x = x_1 f_1 + x_2 f_2$  is equal to  $(x^2) = x_1 x_2$ . The lightlike vectors are located on the coordinate axes; see Fig. 7.11. The timelike vectors comprise the second and fourth quadrants, and the spacelike vectors make up the first and third quadrants.

**Definition 7.50** The set  $V \subset L$  consisting of all lightlike vectors of a pseudo-Euclidean space is called the *light cone* (or *isotropic cone*).

That we call the set  $V$  a *cone* suggests that if it contains some vector  $e$ , then it contains the entire straight line  $\langle e \rangle$ , which follows at once from the definition. The set of timelike vectors is called the *interior* of the cone  $V$ , while the set of spacelike vectors makes up its *exterior*. In the space from Example 7.49, the light cone  $V$  is the union of two straight lines  $\langle f_1 \rangle$  and  $\langle f_2 \rangle$ . A more visual representation of the light cone is given by the following example.

<sup>8</sup>We remark that this terminology differs from what is generally used: Our “spacelike” vectors are usually called “timelike,” and conversely. The difference is explained by the condition  $s = n - 1$  that we have assumed. In the conventional definition of Minkowski space, one usually considers the quadratic form  $-x^2 - y^2 - z^2 + t^2$ , with index of inertia  $s = 1$ , and we need to multiply it by  $-1$  in order that the condition  $s \geq n/2$  be satisfied.

**Fig. 7.12** The light cone

**Example 7.51** We consider the pseudo-Euclidean space  $L$  with  $\dim L = 3$  and index of inertia  $s = 2$ . With the selection of an orthonormal basis  $e_1, e_2, e_3$  such that

$$(e_1^2) = (e_2^2) = 1, \quad (e_3^2) = -1, \quad (e_i, e_j) = 0 \quad \text{for all } i \neq j,$$

the light cone  $V$  is defined by the equation  $x_1^2 + x_2^2 - x_3^2 = 0$ . This is an ordinary right circular cone in three-dimensional space, familiar from a course in analytic geometry; see Fig. 7.12.

We now return to the general case of a pseudo-Euclidean space  $L$  of dimension  $n$  and consider the light cone  $V$  in  $L$  in greater detail. First of all, let us verify that it is “completely circular.” By this we mean the following.

**Lemma 7.52** *Although the cone  $V$  contains along with every vector  $x$  the entire line  $\langle x \rangle$ , it contains no two-dimensional subspace.*

*Proof* Let us assume that  $V$  contains a two-dimensional subspace  $\langle x, y \rangle$ . We choose a vector  $e \in L$  such that  $(e^2) = -1$ . Then the line  $\langle e \rangle$  is a nondegenerate subspace of  $L$ , and we can use the decomposition (7.73):

$$L = \langle e \rangle \oplus \langle e \rangle^\perp. \quad (7.79)$$

From the law of inertia it follows that  $\langle e \rangle^\perp$  is a Euclidean space. Let us apply the decomposition (7.79) to our vectors  $x, y \in V$ . We obtain

$$x = \alpha e + u, \quad y = \beta e + v, \quad (7.80)$$

where  $u$  and  $v$  are vectors in the Euclidean space  $\langle e \rangle^\perp$ , while  $\alpha$  and  $\beta$  are some scalars.

The conditions  $(x^2) = 0$  and  $(y^2) = 0$  can be written as  $\alpha^2 = (u^2)$  and  $\beta^2 = (v^2)$ . Using the same reasoning for the vector  $x + y = (\alpha + \beta)e + u + v$ , which by the assumption  $\langle x, y \rangle \subset V$  is also contained in  $V$ , we obtain the equality

$$(\alpha + \beta)^2 = (u + v, u + v) = (u^2) + 2(u, v) + (v^2) = \alpha^2 + 2(u, v) + \beta^2.$$

Canceling the terms  $\alpha^2$  and  $\beta^2$  on the left- and right-hand sides of the equality, we obtain that  $\alpha\beta = (u, v)$ , that is,  $(u, v)^2 = \alpha^2\beta^2 = (u^2) \cdot (v^2)$ . Thus for the vectors

$\mathbf{u}$  and  $\mathbf{v}$  in the Euclidean space  $\langle \mathbf{e} \rangle^\perp$ , the Cauchy–Schwarz inequality reduces to an equality, from which it follows that  $\mathbf{u}$  and  $\mathbf{v}$  are proportional (see p. 218). Let  $\mathbf{v} = \lambda \mathbf{u}$ . Then the vector  $\mathbf{y} - \lambda \mathbf{x} = (\beta - \lambda \alpha) \mathbf{e}$  is also lightlike. Since  $\langle \mathbf{e}^2 \rangle = -1$ , it follows that  $\beta = \lambda \alpha$ . But then from the relationship (7.80), it follows that  $\mathbf{y} = \lambda \mathbf{x}$ , and this contradicts the assumption  $\dim \langle \mathbf{x}, \mathbf{y} \rangle = 2$ .  $\square$

Let us select an arbitrary timelike vector  $\mathbf{e} \in L$ . Then in the orthogonal complement  $\langle \mathbf{e} \rangle^\perp$  of the line  $\langle \mathbf{e} \rangle$ , the bilinear form  $(\mathbf{x}, \mathbf{y})$  determines a positive definite quadratic form. This implies that  $\langle \mathbf{e} \rangle^\perp \cap V = \{\mathbf{0}\}$ , and the hyperplane  $\langle \mathbf{e} \rangle^\perp$  divides the set  $V \setminus \{\mathbf{0}\}$  into two parts,  $V_+$  and  $V_-$ , consisting of vectors  $\mathbf{x} \in V$  such that in each part, the condition  $(\mathbf{e}, \mathbf{x}) > 0$  or  $(\mathbf{e}, \mathbf{x}) < 0$  is respectively satisfied. We shall call these sets  $V_+$  and  $V_-$  *poles* of the light cone  $V$ . In Fig. 7.12, the plane  $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle$  divides  $V$  into “upper” and “lower” poles  $V_+$  and  $V_-$  for the vector  $\mathbf{e} = \mathbf{e}_3$ .

The partition  $V \setminus \{\mathbf{0}\} = V_+ \cup V_-$  that we have constructed rested on the choice of some timelike vector  $\mathbf{e}$ , and ostensibly, it must depend on it (for example, a change in the vector  $\mathbf{e}$  to  $-\mathbf{e}$  interchanges the poles  $V_+$  and  $V_-$ ). We shall now show that the decomposition  $V \setminus \{\mathbf{0}\} = V_+ \cup V_-$ , without taking into account how we designate each pole, does not depend on the choice of vector  $\mathbf{e}$ , that is, it is a property of the pseudo-Euclidean space itself. To do so, we shall require the following, almost obvious, assertion.

**Lemma 7.53** *Let  $L'$  be a subspace of the pseudo-Euclidean space  $L$  of dimension  $\dim L' \geq 2$ . Then the following statements are equivalent:*

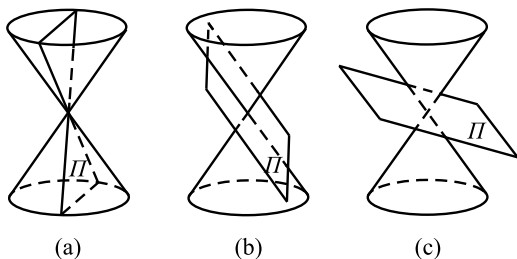
- (1)  $L'$  is a pseudo-Euclidean space.
- (2)  $L'$  contains a timelike vector.
- (3)  $L'$  contains two linearly independent lightlike vectors.

*Proof* If  $L'$  is a pseudo-Euclidean space, then statements (2) and (3) obviously follow from the definition of a pseudo-Euclidean space.

Let us show that statement (2) implies statement (1). Suppose  $L'$  contains a timelike vector  $\mathbf{e}$ . That is,  $\langle \mathbf{e}^2 \rangle < 0$ , whence the subspace  $\langle \mathbf{e} \rangle$  is nondegenerate, and therefore, we have the decomposition (7.79), and moreover, as follows from the law of inertia, the subspace  $\langle \mathbf{e} \rangle^\perp$  is a Euclidean space. If the subspace  $L'$  were degenerate, then there would exist a nonnull vector  $\mathbf{u} \in L'$  such that  $(\mathbf{u}, \mathbf{x}) = 0$  for all  $\mathbf{x} \in L'$ , and in particular, for vectors  $\mathbf{e}$  and  $\mathbf{u}$ . The condition  $(\mathbf{u}, \mathbf{e}) = 0$  implies that the vector  $\mathbf{u}$  is contained in  $\langle \mathbf{e} \rangle^\perp$ , while the condition  $(\mathbf{u}, \mathbf{u}) = 0$  implies that the vector  $\mathbf{u}$  is lightlike. But this is impossible, since the subspace  $\langle \mathbf{e} \rangle^\perp$  is a Euclidean space and cannot contain lightlike vectors. This contradiction shows that the subspace  $L'$  is nondegenerate, and therefore, it exhibits the decomposition (7.73). Taking into account the law of inertia, it follows from this that the subspace  $L'$  is a pseudo-Euclidean space.

Let us show that statement (3) implies statement (1). Suppose the subspace  $L'$  contains linearly independent lightlike vectors  $\mathbf{f}_1$  and  $\mathbf{f}_2$ . We shall show that the plane  $\Pi = \langle \mathbf{f}_1, \mathbf{f}_2 \rangle$  contains a timelike vector  $\mathbf{e}$ . Then obviously,  $\mathbf{e}$  is contained

**Fig. 7.13** The plane  $\Pi$  in a three-dimensional pseudo-Euclidean space



in  $L'$ , and by what was proved above, the subspace  $L'$  is a pseudo-Euclidean space. Every vector  $e \in \Pi$  can be represented in the form  $e = \alpha f_1 + \beta f_2$ . From this, we obtain  $(e^2) = 2\alpha\beta(f_1, f_2)$ . We note that  $(f_1, f_2) \neq 0$ , since in the contrary case, for each vector  $e \in \Pi$ , the equality  $(e^2) = 0$  would be satisfied, implying that the plane  $\Pi$  lies completely in the light cone  $V$ , which contradicts Lemma 7.52. Thus  $(f_1, f_2) \neq 0$ , and choosing coordinates  $\alpha$  and  $\beta$  such that the sign of their product is opposite to the sign of  $(f_1, f_2)$ , we obtain the vector  $e$ , for which  $(e^2) < 0$ .  $\square$

**Example 7.54** Let us consider the three-dimensional pseudo-Euclidean space  $L$  from Example 7.51 and a plane  $\Pi$  in  $L$ . The property of a plane  $\Pi$  being a Euclidean space, a pseudo-Euclidean space, or degenerate is clearly illustrated in Fig. 7.13.

In Fig. 7.13(a), the plane  $\Pi$  intersects the light cone  $V$  in two lines, corresponding to two linearly independent lightlike vectors. Clearly, this is equivalent to the condition that  $\Pi$  also intersects the interior of the light cone, which consists of timelike vectors, and therefore is a pseudo-Euclidean plane. In Fig. 7.13(c), it is shown that the plane  $\Pi$  intersects  $V$  only in its vertex, that is,  $\Pi \cap V = \{0\}$ . This implies that the plane  $\Pi$  is a Euclidean space, since every nonnull vector  $e \in \Pi$  lies outside the cone  $V$ , that is,  $(e^2) > 0$ .

Finally, in Fig. 7.13(b) is shown the intermediate variant: the plane  $\Pi$  intersects the cone  $V$  in a single line, that is, it is tangent to it. Since the plane  $\Pi$  contains lightlike vectors (lying on this line), it follows that it cannot be a Euclidean space, and since it does not contain timelike vectors, it follows by Lemma 7.53 that it cannot be a pseudo-Euclidean space. This implies that  $\Pi$  is degenerate.

This is not difficult to verify in another way if we write down the matrix of the restriction of the inner product to the plane  $\Pi$ . Suppose that in the orthonormal basis  $e_1, e_2, e_3$  from Example 7.49, this plane is defined by the equation  $x_3 = \alpha x_1 + \beta x_2$ . Then the vectors  $g_1 = e_1 + \alpha e_3$  and  $g_2 = e_2 + \beta e_3$  form a basis of  $\Pi$  in which the restriction of the inner product has matrix  $\begin{pmatrix} 1-\alpha^2 & -\alpha\beta \\ -\alpha\beta & 1-\beta^2 \end{pmatrix}$  with determinant  $\Delta = (1 - \alpha^2)(1 - \beta^2) - (\alpha\beta)^2$ . On the other hand, the assumption of tangency of the plane  $\Pi$  and cone  $V$  amounts to the discriminant of the quadratic form  $x_1^2 + x_2^2 - (\alpha x_1 + \beta x_2)^2$  in the variables  $x_1$  and  $x_2$  being equal to zero. It is easily verified that this discriminant is equal to  $-\Delta$ , and this implies that it is zero precisely when the determinant of this matrix is zero.



**Theorem 7.55** *The partition of the light cone  $V$  into two poles  $V_+$  and  $V_-$  does not depend on the choice of timelike vector  $\mathbf{e}$ . In particular, the linearly independent lightlike vectors  $\mathbf{x}$  and  $\mathbf{y}$  lie in one pole if and only if  $\langle \mathbf{x}, \mathbf{y} \rangle < 0$ .*

*Proof* Let us assume that for some choice of timelike vector  $\mathbf{e}$ , the lightlike vectors  $\mathbf{x}$  and  $\mathbf{y}$  lie in one pole of the light cone  $V$ , and let us show that then, for any choice  $\mathbf{e}$ , they will always belong to the same pole. The case that the vectors  $\mathbf{x}$  and  $\mathbf{y}$  are proportional, that is,  $\mathbf{y} = \lambda \mathbf{x}$ , is obvious. Indeed, since  $\langle \mathbf{e} \rangle^\perp \cap V = \{\mathbf{0}\}$ , it follows that  $\langle \mathbf{e}, \mathbf{x} \rangle \neq 0$ , and this implies that the vectors  $\mathbf{x}$  and  $\mathbf{y}$  belong to one pole if and only if  $\lambda > 0$ , independent of the choice of the vector  $\mathbf{e}$ .

Now let us consider the case that  $\mathbf{x}$  and  $\mathbf{y}$  are linearly independent. Then  $\langle \mathbf{x}, \mathbf{y} \rangle \neq 0$ , since otherwise, the entire plane  $\langle \mathbf{x}, \mathbf{y} \rangle$  would be contained in the light cone  $V$ , which by Lemma 7.52, is impossible. Let us prove that regardless of what timelike vector  $\mathbf{e}$  we have chosen for the partition  $V \setminus \{\mathbf{0}\} = V_+ \cup V_-$ , the vectors  $\mathbf{x}, \mathbf{y} \in V \setminus \{\mathbf{0}\}$  belong to one pole if and only if  $\langle \mathbf{x}, \mathbf{y} \rangle < 0$ . Let us note that this question, strictly speaking, relates not to the entire space  $L$ , but only to the subspace  $\langle \mathbf{e}, \mathbf{x}, \mathbf{y} \rangle$ , whose dimension, by the assumptions we have made, is equal to 2 or 3, depending on whether the vector  $\mathbf{e}$  does or does not lie in the plane  $\langle \mathbf{x}, \mathbf{y} \rangle$ .

Let us consider first the case  $\dim \langle \mathbf{e}, \mathbf{x}, \mathbf{y} \rangle = 2$ , that is,  $\mathbf{e} \in \langle \mathbf{x}, \mathbf{y} \rangle$ . Then let us set  $\mathbf{e} = \alpha \mathbf{x} + \beta \mathbf{y}$ . Consequently,  $\langle \mathbf{e}, \mathbf{x} \rangle = \beta \langle \mathbf{x}, \mathbf{y} \rangle$  and  $\langle \mathbf{e}, \mathbf{y} \rangle = \alpha \langle \mathbf{x}, \mathbf{y} \rangle$ , since  $\mathbf{x}, \mathbf{y} \in V$ . By definition, vectors  $\mathbf{x}$  and  $\mathbf{y}$  are in the same pole if and only if  $\langle \mathbf{e}, \mathbf{x} \rangle \langle \mathbf{e}, \mathbf{y} \rangle > 0$ . But since  $\langle \mathbf{e}, \mathbf{x} \rangle \langle \mathbf{e}, \mathbf{y} \rangle = \alpha \beta \langle \mathbf{x}, \mathbf{y} \rangle^2$ , this condition is equivalent to the inequality  $\alpha \beta > 0$ . The vector  $\mathbf{e}$  is timelike, and therefore,  $\langle \mathbf{e}^2 \rangle < 0$ , and in view of the equality  $\langle \mathbf{e}^2 \rangle = 2\alpha\beta \langle \mathbf{x}, \mathbf{y} \rangle$ , we obtain that the condition  $\alpha\beta > 0$  is equivalent to  $\langle \mathbf{x}, \mathbf{y} \rangle < 0$ .

Let us now consider the case that  $\dim \langle \mathbf{e}, \mathbf{x}, \mathbf{y} \rangle = 3$ . The space  $\langle \mathbf{e}, \mathbf{x}, \mathbf{y} \rangle$  contains the timelike vector  $\mathbf{e}$ . Consequently, by Lemma 7.53, it is a pseudo-Euclidean space, and its subspace  $\langle \mathbf{x}, \mathbf{y} \rangle$  is nondegenerate, since  $\langle \mathbf{x}, \mathbf{y} \rangle \neq 0$  and  $\langle \mathbf{x}^2 \rangle = \langle \mathbf{y}^2 \rangle = 0$ . Thus here the decomposition (7.73) takes the form

$$\langle \mathbf{e}, \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle \oplus \langle \mathbf{h} \rangle, \quad (7.81)$$

where the space  $\langle \mathbf{h} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle^\perp$  is one-dimensional. On the left-hand side of the decomposition (7.81) stands a three-dimensional pseudo-Euclidean space, and the space  $\langle \mathbf{x}, \mathbf{y} \rangle$  is a two-dimensional pseudo-Euclidean space; therefore, by the law of inertia, the space  $\langle \mathbf{h} \rangle$  is a Euclidean space. Thus for the vector  $\mathbf{e}$ , we have the representation

$$\mathbf{e} = \alpha \mathbf{x} + \beta \mathbf{y} + \gamma \mathbf{h}, \quad \langle \mathbf{h}, \mathbf{x} \rangle = 0, \quad \langle \mathbf{h}, \mathbf{y} \rangle = 0.$$

From this follows the equality

$$\langle \mathbf{e}, \mathbf{x} \rangle = \beta \langle \mathbf{x}, \mathbf{y} \rangle, \quad \langle \mathbf{e}, \mathbf{y} \rangle = \alpha \langle \mathbf{x}, \mathbf{y} \rangle, \quad \langle \mathbf{e}^2 \rangle = 2\alpha\beta \langle \mathbf{x}, \mathbf{y} \rangle + \gamma^2 \langle \mathbf{h}^2 \rangle.$$

Taking into account the fact that  $\langle \mathbf{e}^2 \rangle < 0$  and  $\langle \mathbf{h}^2 \rangle > 0$ , from the last of these relationships, we obtain that  $\alpha\beta \langle \mathbf{x}, \mathbf{y} \rangle < 0$ . The condition that the vectors  $\mathbf{x}$  and  $\mathbf{y}$  lie in one pole can be expressed as the inequality  $\langle \mathbf{e}, \mathbf{x} \rangle \langle \mathbf{e}, \mathbf{y} \rangle > 0$ , that is,  $\alpha\beta > 0$ .

Since  $\alpha\beta(\mathbf{x}, \mathbf{y}) < 0$ , it follows as in the previous case that this is equivalent to the condition  $(\mathbf{x}, \mathbf{y}) < 0$ .  $\square$

*Remark 7.56* As we did in Sect. 3.2 in connection with the partition of a vector space  $L$  by a hyperplane  $L'$ , it is possible to ascertain that the partition of the set  $V \setminus \mathbf{0}$  coincides with its partition into two path-connected components  $V_+$  and  $V_-$ . From this we can obtain another proof of Theorem 7.55 without using any formulas.

A pseudo-Euclidean space emerges in the following remarkable relationship.

**Theorem 7.57** *For every pair of timelike vectors  $\mathbf{x}$  and  $\mathbf{y}$ , the reverse of the Cauchy–Schwarz inequality is satisfied:*

$$(\mathbf{x}, \mathbf{y})^2 \geq (\mathbf{x}^2) \cdot (\mathbf{y}^2), \quad (7.82)$$

which reduces to an equality if and only if  $\mathbf{x}$  and  $\mathbf{y}$  are proportional.

*Proof* Let us consider the subspace  $\langle \mathbf{x}, \mathbf{y} \rangle$ , in which are contained all the vectors of interest to us. If the vectors  $\mathbf{x}$  and  $\mathbf{y}$  are proportional, that is,  $\mathbf{y} = \lambda \mathbf{x}$ , where  $\lambda$  is some scalar, then the inequality (7.82) obviously reduces to a tautological equality. Thus we may assume that  $\dim \langle \mathbf{x}, \mathbf{y} \rangle = 2$ , that is, we may suppose ourselves to be in the situation considered in Example 7.49.

As we have seen, in the space  $\langle \mathbf{x}, \mathbf{y} \rangle$ , there exists a basis  $\mathbf{f}_1, \mathbf{f}_2$  for which the relationship  $(\mathbf{f}_1^2) = (\mathbf{f}_2^2) = 0$ ,  $(\mathbf{f}_1, \mathbf{f}_2) = \frac{1}{2}$  holds. Writing the vectors  $\mathbf{x}$  and  $\mathbf{y}$  in this basis, we obtain the expressions

$$\mathbf{x} = x_1 \mathbf{f}_1 + x_2 \mathbf{f}_2, \quad \mathbf{y} = y_1 \mathbf{f}_1 + y_2 \mathbf{f}_2,$$

from which it follows that

$$(\mathbf{x}^2) = x_1 x_2, \quad (\mathbf{y}^2) = y_1 y_2, \quad (\mathbf{x}, \mathbf{y}) = \frac{1}{2}(x_1 y_2 + x_2 y_1).$$

Substituting these expressions into (7.82), we see that we have to verify the inequality  $(x_1 y_2 + x_2 y_1)^2 \geq 4x_1 x_2 y_1 y_2$ . Having carried out in the last inequality the obvious transformations, we see that this is equivalent to the inequality

$$(x_1 y_2 - x_2 y_1)^2 \geq 0, \quad (7.83)$$

which holds for all real values of the variables. Moreover, it is obvious that the inequality (7.83) reduces to an equality if and only if  $x_1 y_2 - x_2 y_1 = 0$ , that is, if and only if the determinant  $\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix}$  equals 0, and this implies that the vectors  $\mathbf{x} = (x_1, x_2)$  and  $\mathbf{y} = (y_1, y_2)$  are proportional.  $\square$

From Theorem 7.57 we obtain the following useful corollary.

**Corollary 7.58** *Two timelike vectors  $\mathbf{x}$  and  $\mathbf{y}$  cannot be orthogonal.*

*Proof* Indeed, if  $(x, y) = 0$ , then from the inequality (7.82), it follows that  $(x^2) \cdot (y^2) \leq 0$ , and this contradicts the condition  $(x^2) < 0$  and  $(y^2) < 0$ .  $\square$

Similar to the partition of the light cone  $V$  into two poles, we can also partition its interior into two parts. Namely, we shall say that timelike vectors  $e$  and  $e'$  lie *inside* one pole of the light cone  $V$  if the inner products  $(e, x)$  and  $(e', x)$  have the same sign for all vectors  $x \in V$  and lie *inside* different poles if these inner products have opposite signs.

A set  $M \subset L$  is said to be convex if for every pair of vectors  $e, e' \in M$ , the vectors  $g_t = te + (1 - t)e'$  are also in  $M$  for all  $t \in [0, 1]$ . We shall prove that the interior of each pole of the light cone  $V$  is convex, that is, the vector  $g_t$  lies in the same pole as  $e$  and  $e'$  for all  $t \in [0, 1]$ . To this end, let us note that in the expression  $(g_t, x) = t(e, x) + (1 - t)(e', x)$ , the coefficients  $t$  and  $1 - t$  are nonnegative, and the inner products  $(e, x)$  and  $(e', x)$  have the same sign. Therefore, for every vector  $x \in V$ , the inner product  $(g_t, x)$  has the same sign as  $(e, x)$  and  $(e', x)$ .

**Lemma 7.59** *Timelike vectors  $e$  and  $e'$  lie inside one pole of the light cone  $V$  if and only if  $(e, e') < 0$ .*

*Proof* If timelike vectors  $e$  and  $e'$  lie inside one pole, then by definition, we have the inequality  $(e, x) \cdot (e', x) > 0$  for all  $x \in V$ . Let us assume that  $(e, e') \geq 0$ . As we established above, the vector  $g_t = te + (1 - t)e'$  is timelike and lies inside the same pole as  $e$  and  $e'$  for all  $t \in [0, 1]$ .

Let us consider the inner product  $(g_t, e) = t(e, e) + (1 - t)(e, e')$  as a function of the variable  $t \in [0, 1]$ . It is obvious that this function is continuous and that it assumes for  $t = 0$  the value  $(e, e') \geq 0$ , and for  $t = 1$  the value  $(e, e) < 0$ . Therefore, as is proved in a course in calculus, there exists a value  $\tau \in [0, 1]$  such that  $(g_\tau, e) = 0$ . But this contradicts Corollary 7.58.

Thus we have proved that if vectors  $e$  and  $e'$  lie inside one pole of the cone  $V$ , then  $(e, e') < 0$ . The converse assertion is obvious. Let  $e$  and  $e'$  lie inside different poles, for instance,  $e$  is within  $V_+$ , while  $e'$  is within  $V_-$ . Then we have by definition that the vector  $-e'$  lies inside the pole  $V_+$ , and therefore,  $(e, -e') < 0$ , that is,  $(e, e') > 0$ .  $\square$

## 7.8 Lorentz Transformations

In this section, we shall examine an analogue of orthogonal transformations for pseudo-Euclidean spaces called *Lorentz* transformations. Such transformations have numerous applications in physics.<sup>9</sup> They are also defined by the condition of preserving the inner product.

<sup>9</sup>For example, a Lorentz transformation of Minkowski space—a four-dimensional pseudo-Euclidean space—plays the same role in the special theory of relativity (which is where the term Lorentz transformation comes from) as that played by the Galilean transformations, which describe the passage from one inertial reference frame to another in classical Newtonian mechanics.

**Definition 7.60** A linear transformation  $\mathcal{U}$  of a pseudo-Euclidean space  $L$  is called a *Lorentz transformation* if the relationship

$$(\mathcal{U}(\mathbf{x}), \mathcal{U}(\mathbf{y})) = (\mathbf{x}, \mathbf{y}) \quad (7.84)$$

is satisfied for all vectors  $\mathbf{x}, \mathbf{y} \in L$ .

As in the case of orthogonal transformations, it suffices that the condition (7.84) be satisfied for all vectors  $\mathbf{x} = \mathbf{y}$  of the pseudo-Euclidean space  $L$ . The proof of this coincides completely with the proof of the analogous assertion in Sect. 7.2.

Here, as in the case of Euclidean spaces, we shall make use of the inner product  $(\mathbf{x}, \mathbf{y})$  in order to identify  $L^*$  with  $L$  (let us recall that for this, we need only the nonsingularity of the bilinear form  $(\mathbf{x}, \mathbf{y})$  and not the positive definiteness of the associated quadratic form  $(\mathbf{x}^2)$ ). As a result, for an arbitrary linear transformation  $\mathcal{A} : L \rightarrow L$ , we may consider  $\mathcal{A}^*$  also as a transformation of the space  $L$  into itself. Repeating the same arguments that we employed in the case of Euclidean spaces, we obtain that  $|\mathcal{A}^*| = |\mathcal{A}|$ . In particular, from definition (7.84), it follows that for a Lorentz transformation  $\mathcal{U}$ , we have the relationship

$$U^*AU = A, \quad (7.85)$$

where  $U$  is the matrix of the transformation  $\mathcal{U}$  in an arbitrary basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the space  $L$ , and  $A = (a_{ij})$  is the Gram matrix of the bilinear form  $(\mathbf{x}, \mathbf{y})$ , that is, the matrix with elements  $a_{ij} = (\mathbf{e}_i, \mathbf{e}_j)$ .

The bilinear form  $(\mathbf{x}, \mathbf{y})$  is nonsingular, that is,  $|A| \neq 0$ , and from the relationship (7.85) follows the equality  $|\mathcal{U}|^2 = 1$ , from which we obtain that  $|\mathcal{U}| = \pm 1$ . As in the case of a Euclidean space, a transformation with determinant equal to 1 is called *proper*, while if the determinant is equal to  $-1$ , it is *improper*.

It follows from the definition that every Lorentz transformation maps the light cone  $V$  into itself. It follows from Theorem 7.55 that a Lorentz transformation either maps each pole into itself (that is,  $\mathcal{U}(V_+) = V_+$  and  $\mathcal{U}(V_-) = V_-$ ), or else interchanges them (that is,  $\mathcal{U}(V_+) = V_-$  and  $\mathcal{U}(V_-) = V_+$ ). Let us associate with each Lorentz transformation  $\mathcal{U}$  the number  $\nu(\mathcal{U}) = +1$  in the first case, and  $\nu(\mathcal{U}) = -1$  in the second. Like the determinant  $|\mathcal{U}|$ , this number  $\nu(\mathcal{U})$  is a natural characteristic of the associated Lorentz transformation. Let us denote the pair of numbers  $(|\mathcal{U}|, \nu(\mathcal{U}))$  by  $\varepsilon(\mathcal{U})$ . It is obvious that

$$\varepsilon(\mathcal{U}^{-1}) = \varepsilon(\mathcal{U}), \quad \varepsilon(\mathcal{U}_1\mathcal{U}_2) = \varepsilon(\mathcal{U}_1)\varepsilon(\mathcal{U}_2),$$

where on the right-hand side, it is understood that the first and second components of the pairs are multiplied separately. We shall soon see that in an arbitrary pseudo-Euclidean space, there exist Lorentz transformations  $\mathcal{U}$  of all four types, that is, with  $\varepsilon(\mathcal{U})$  taking all values

$$(+1, +1), \quad (+1, -1), \quad (-1, +1), \quad (-1, -1).$$

This property is sometimes interpreted as saying that a pseudo-Euclidean space has not two (as in the case of a Euclidean space), but *four* orientations.

Like orthogonal transformations of a Euclidean space, Lorentz transformations are characterized by the fact that they map an orthonormal basis of a pseudo-Euclidean space to an orthonormal basis. Indeed, suppose that for the vectors of the orthonormal basis  $e_1, \dots, e_n$ , the equalities

$$(e_i, e_j) = 0 \quad \text{for } i \neq j, \quad (e_1^2) = \dots = (e_{n-1}^2) = 1, \quad (e_n^2) = -1 \quad (7.86)$$

are satisfied. Then from the condition (7.84), it follows that the images  $\mathcal{U}(e_1), \dots, \mathcal{U}(e_n)$  satisfy analogous equalities, that is, they form an orthonormal basis in  $L$ . Conversely, if for the vectors  $e_i$ , the equality (7.86) is satisfied and analogous equalities hold for the vectors  $\mathcal{U}(e_i)$ , then as is easily verified, for arbitrary vectors  $x$  and  $y$  of the pseudo-Euclidean space  $L$ , the relationship (7.84) is satisfied.

Two orthonormal bases are said to have the *same orientation* if for a Lorentz transformation  $\mathcal{U}$  taking one basis to the other,  $\varepsilon(\mathcal{U}) = (+1, +1)$ . The choice of a class of bases with the same orientation is called an *orientation* of the pseudo-Euclidean space  $L$ . Taking for now on faith the fact (which will be proved a little later) that there exist Lorentz transformations  $\mathcal{U}$  with all theoretically possible  $\varepsilon(\mathcal{U})$ , we see that in a pseudo-Euclidean space, it is possible to introduce exactly four orientations.

*Example 7.61* Let us consider some concepts about pseudo-Euclidean spaces that we encountered in Example 7.49, that is, for  $\dim L = 2$  and  $s = 1$ . As we have seen, in this space, there exists a basis  $f_1, f_2$  for which the relationships  $(f_1^2) = (f_2^2) = 0$ ,  $(f_1, f_2) = \frac{1}{2}$ , are satisfied, and the scalar square of the vector  $x = xf_1 + yf_2$  is equal to  $(x^2) = xy$ . If  $\mathcal{U} : L \rightarrow L$  is a Lorentz transformation given by the formula

$$x' = ax + by, \quad y' = cx + dy,$$

then the equality  $(\mathcal{U}(x), \mathcal{U}(x)) = (x, x)$  for the vector  $x = xf_1 + yf_2$  takes the form  $x'y' = xy$ , that is,  $(ax + by)(cx + dy) = xy$  for all  $x$  and  $y$ . From this, we obtain

$$ac = 0, \quad bd = 0, \quad ad + bc = 1.$$

In view of the equality  $ad + bc = 1$ , the values  $a = b = 0$  are impossible.

If  $a \neq 0$ , then  $c = 0$ , and this implies that  $ad = 1$ , that is,  $d = a^{-1} \neq 0$  and  $b = 0$ . Thus the transformation  $\mathcal{U}$  has the form

$$x' = ax, \quad y' = a^{-1}y. \quad (7.87)$$

This is a proper transformation.

On the other hand, if  $b \neq 0$ , then  $d = 0$ , and this implies that  $c = b^{-1}$ ,  $a = 0$ . The transformation  $\mathcal{U}$  has in this case the form

$$x' = by, \quad y' = b^{-1}x. \quad (7.88)$$

This is an improper transformation.

If we write the transformation  $\mathcal{U}$  in the form (7.87) or (7.88), depending on whether it is proper or improper, then the sign of the number  $a$  or respectively  $b$  indicates whether  $\mathcal{U}$  interchanges the poles of the light cone or preserves each of them. Namely, let us prove that the transformation (7.87) causes the poles to change places if  $a < 0$ , and preserves them if  $a > 0$ . And analogously, the transformation (7.88) interchanges the poles if  $b < 0$  and preserves them if  $b > 0$ .

By Theorem 7.55, the partition of the light cone  $V$  into two poles  $V_+$  and  $V_-$  does not depend on the choice of timelike vector, and therefore, by Lemma 7.59, we need only determine the sign of the inner product  $(\mathbf{e}, \mathcal{U}(\mathbf{e}))$  for an arbitrary timelike vector  $\mathbf{e}$ . Let  $\mathbf{e} = x\mathbf{f}_1 + y\mathbf{f}_2$ . Then  $(\mathbf{e}^2) = xy < 0$ . In the case that  $\mathcal{U}$  is a proper transformation, we have formula (7.87), from which it follows that

$$\mathcal{U}(\mathbf{e}) = ax\mathbf{f}_1 + a^{-1}y\mathbf{f}_2, \quad (\mathbf{e}, \mathcal{U}(\mathbf{e})) = (a + a^{-1})xy.$$

Since  $xy < 0$ , the inner product  $(\mathbf{e}, \mathcal{U}(\mathbf{e}))$  is negative if  $a + a^{-1} > 0$ , and positive if  $a + a^{-1} < 0$ . But it is obvious that  $a + a^{-1} > 0$  for  $a > 0$ , and  $a + a^{-1} < 0$  for  $a < 0$ . Thus for  $a > 0$ , we have  $(\mathbf{e}, \mathcal{U}(\mathbf{e})) < 0$ , and by Lemma 7.59, the vectors  $\mathbf{e}$  and  $\mathcal{U}(\mathbf{e})$  lie inside one pole. Consequently, the transformation  $\mathcal{U}$  preserves the poles of the light cone. Analogously, for  $a < 0$ , we obtain  $(\mathbf{e}, \mathcal{U}(\mathbf{e})) > 0$ , that is,  $\mathbf{e}$  and  $\mathcal{U}(\mathbf{e})$  lie inside different poles, and therefore, the transformation  $\mathcal{U}$  interchanges the poles.

The case of an improper transformation can be examined with the help of formula (7.88). Reasoning analogously to what has gone before, we obtain from it the relationships

$$\mathcal{U}(\mathbf{e}) = b^{-1}y\mathbf{f}_1 + bx\mathbf{f}_2, \quad (\mathbf{e}, \mathcal{U}(\mathbf{e})) = bx^2 + b^{-1}y^2,$$

from which it is clear that now the sign of  $(\mathbf{e}, \mathcal{U}(\mathbf{e}))$  coincides with the sign of the number  $b$ .

*Example 7.62* It is sometimes convenient to use the fact that a Lorentz transformation of a pseudo-Euclidean plane can be written in an alternative form, using the hyperbolic sine and cosine. We saw earlier (formulas (7.87) and (7.88)) that in the basis  $\mathbf{f}_1, \mathbf{f}_2$  defined by the relationship (7.78), proper and improper Lorentz transformations are given respectively by the equalities

$$\begin{aligned} \mathcal{U}(\mathbf{f}_1) &= a\mathbf{f}_1, & \mathcal{U}(\mathbf{f}_2) &= a^{-1}\mathbf{f}_2, \\ \mathcal{U}(\mathbf{f}_1) &= b\mathbf{f}_2, & \mathcal{U}(\mathbf{f}_2) &= b^{-1}\mathbf{f}_1. \end{aligned}$$

From this, it is not difficult to derive that in the orthonormal basis  $\mathbf{e}_1, \mathbf{e}_2$ , related to  $\mathbf{f}_1, \mathbf{f}_2$  by formula (7.78), these transformations are given respectively by the equalities

$$\begin{aligned} \mathcal{U}(\mathbf{e}_1) &= \frac{a + a^{-1}}{2}\mathbf{e}_1 + \frac{a - a^{-1}}{2}\mathbf{e}_2, \\ \mathcal{U}(\mathbf{e}_2) &= \frac{a - a^{-1}}{2}\mathbf{e}_1 + \frac{a + a^{-1}}{2}\mathbf{e}_2, \end{aligned} \tag{7.89}$$

$$\begin{aligned}
\mathcal{U}(e_1) &= \frac{b+b^{-1}}{2}e_1 - \frac{b-b^{-1}}{2}e_2, \\
\mathcal{U}(e_2) &= \frac{b-b^{-1}}{2}e_1 + \frac{b+b^{-1}}{2}e_2.
\end{aligned} \tag{7.90}$$

Setting here  $a = \pm e^\psi$  and  $b = \pm e^\psi$ , where the sign  $\pm$  coincides with the sign of the number  $a$  or  $b$  in formula (7.89) or (7.90) respectively, we obtain that the matrices of the proper transformations have the form

$$\begin{pmatrix} \cosh \psi & \sinh \psi \\ \sinh \psi & \cosh \psi \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} -\cosh \psi & -\sinh \psi \\ -\sinh \psi & -\cosh \psi \end{pmatrix}, \tag{7.91}$$

while the matrices of the improper transformations have the form

$$\begin{pmatrix} \cosh \psi & \sinh \psi \\ -\sinh \psi & -\cosh \psi \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} -\cosh \psi & -\sinh \psi \\ \sinh \psi & \cosh \psi \end{pmatrix}, \tag{7.92}$$

where  $\sinh \psi = (e^\psi - e^{-\psi})/2$  and  $\cosh \psi = (e^\psi + e^{-\psi})/2$  are the hyperbolic sine and cosine.

**Theorem 7.63** *In every pseudo-Euclidean space there exist Lorentz transformations  $\mathcal{U}$  with all four possible values of  $\varepsilon(\mathcal{U})$ .*

*Proof* For the case  $\dim L = 2$ , we have already proved the theorem: In Example 7.62, we saw that there exist four distinct types of Lorentz transformation of a pseudo-Euclidean space having in a suitable orthonormal basis the matrices (7.91), (7.92). It is obvious that with these matrices, the transformation  $\mathcal{U}$  gives all possible values  $\varepsilon(\mathcal{U})$ .

Let us now move on to the general case  $\dim L > 2$ . Let us choose in the pseudo-Euclidean space  $L$  an arbitrary timelike vector  $e$  and any  $e'$  not proportional to it. By Lemma 7.53, the two-dimensional space  $\langle e, e' \rangle$  is a pseudo-Euclidean space (therefore nondegenerate), and we have the decomposition

$$L = \langle e, e' \rangle \oplus \langle e, e' \rangle^\perp.$$

From the law of inertia, it follows that the space  $\langle e, e' \rangle^\perp$  is a Euclidean space. In Example 7.62, we saw that in the pseudo-Euclidean plane  $\langle e, e' \rangle$ , there exists a Lorentz transformation  $\mathcal{U}_1$  with arbitrary value  $\varepsilon(\mathcal{U}_1)$ . Let us define the transformation  $\mathcal{U} : L \rightarrow L$  as  $\mathcal{U}_1$  in  $\langle e, e' \rangle$  and  $\mathcal{E}$  in  $\langle e, e' \rangle^\perp$ , that is, for a vector  $x = y + z$ , where  $y \in \langle e, e' \rangle$  and  $z \in \langle e, e' \rangle^\perp$ , we shall set  $\mathcal{U}(x) = \mathcal{U}_1(y) + z$ . Then  $\mathcal{U}$  is clearly a Lorentz transformation, and  $\varepsilon(\mathcal{U}) = \varepsilon(\mathcal{U}_1)$ .  $\square$

There is an analogue to Theorem 7.24 for Lorentz transformations.

**Theorem 7.64** *If a space  $L'$  is invariant with respect to a Lorentz transformation  $\mathcal{U}$ , then its orthogonal complement  $(L')^\perp$  is also invariant with respect to  $\mathcal{U}$ .*

*Proof* The proof of this theorem is an exact repetition of the proof of Theorem 7.24, since there, we did not use the positive definiteness of the quadratic form  $(\mathbf{x}^2)$  associated with the bilinear form  $(\mathbf{x}, \mathbf{y})$ , but only its nonsingularity. See Remark 7.25 on p. 227.  $\square$

The study of a Lorentz transformation of a pseudo-Euclidean space is reduced to the analogous question for orthogonal transformations of a Euclidean space, based on the following result.

**Theorem 7.65** *For every Lorentz transformation  $\mathcal{U}$  of a pseudo-Euclidean space  $L$ , there exist nondegenerate subspaces  $L_0$  and  $L_1$  invariant with respect to  $\mathcal{U}$  such that  $L$  has the orthogonal decomposition*

$$L = L_0 \oplus L_1, \quad L_0 \perp L_1, \quad (7.93)$$

where the subspace  $L_0$  is a Euclidean space, and the dimension of  $L_1$  is equal to 1, 2, or 3.

It follows from the law of inertia that if  $\dim L_1 = 1$ , then  $L_1$  is spanned by a timelike vector. If  $\dim L_1 = 2$  or 3, then the pseudo-Euclidean space  $L_1$  can be represented in turn by a direct sum of subspaces of lower dimension invariant with respect to  $\mathcal{U}$ . However, such a decomposition is no longer necessarily orthogonal (see Example 7.48).

*Proof of Theorem 7.65* The proof is by induction on  $n$ , the dimension of the space  $L$ . For  $n = 2$ , the assertion of the theorem is obvious—in the decomposition (7.93) one has only to set  $L_0 = (\mathbf{0})$  and  $L_1 = L$ .<sup>10</sup>

Now let  $n > 2$ , and suppose that the assertion of the theorem has been proved for all pseudo-Euclidean spaces of dimension less than  $n$ . We shall use results obtained in Chaps. 4 and 5 on linear transformations of a vector space into itself. Obviously, one of the following three cases must hold: the transformation  $\mathcal{U}$  has a complex eigenvalue,  $\mathcal{U}$  has two linearly independent eigenvectors, or the space  $L$  is cyclic for  $\mathcal{U}$ , corresponding to the only real eigenvalue. Let us consider the three cases separately.

*Case 1.* A linear transformation  $\mathcal{U}$  of a real vector space  $L$  has a complex eigenvalue  $\lambda$ . As established in Sect. 4.3, then  $\mathcal{U}$  also has the complex conjugate eigenvalue  $\bar{\lambda}$ , and moreover, to the pair  $\lambda, \bar{\lambda}$  there corresponds the two-dimensional real invariant subspace  $L' \subset L$ , which contains no real eigenvectors. It is obvious that  $L'$  cannot be a pseudo-Euclidean space: for then the restriction of  $\mathcal{U}$  to  $L'$  would have real eigenvalues, and  $L'$  would contain real eigenvectors of the transformation  $\mathcal{U}$ ; see Examples 7.61 and 7.62. Let us show that  $L'$  is nondegenerate.

<sup>10</sup>The nondegeneracy of the subspace  $L_0 = (\mathbf{0})$  relative to a bilinear form follows from the definitions given on pages 266 and 195. Indeed, the rank of the restriction of the bilinear form to the subspace  $(\mathbf{0})$  is zero, and therefore, it coincides with  $\dim(\mathbf{0})$ .



Suppose that  $L'$  is degenerate. Then it contains a lightlike vector  $e \neq 0$ . Since  $\mathcal{U}$  is a Lorentz transformation, the vector  $\mathcal{U}(e)$  is also lightlike, and since the subspace  $L'$  is invariant with respect to  $\mathcal{U}$ , it follows that  $\mathcal{U}(e)$  is contained in  $L'$ . Therefore, the subspace  $L'$  contains two lightlike vectors:  $e$  and  $\mathcal{U}(e)$ . By Lemma 7.53, these vectors cannot be linearly independent, since then  $L'$  would be a pseudo-Euclidean space, but that would contradict our assumption that  $L'$  is degenerate. From this, it follows that the vector  $\mathcal{U}(e)$  is proportional to  $e$ , and that implies that  $e$  is an eigenvector of the transformation  $\mathcal{U}$ , which, as we have observed above, cannot be. This contradiction means that the subspace  $L'$  is nondegenerate, and as a consequence, it is a Euclidean space.

*Case 2.* The linear transformation  $\mathcal{U}$  has two linearly independent eigenvectors:  $e_1$  and  $e_2$ . If at least one of them is not lightlike, that is,  $(e_i^2) \neq 0$ , then  $L' = \langle e_i \rangle$  is a nondegenerate invariant subspace of dimension 1. And if both eigenvectors  $e_1$  and  $e_2$  are lightlike, then by Lemma 7.53, the subspace  $L' = \langle e_1, e_2 \rangle$  is an invariant pseudo-Euclidean plane.

Thus in both cases, the transformation  $\mathcal{U}$  has a nondegenerate invariant subspace  $L'$  of dimension 1 or 2. This means that in both cases, we have an orthogonal decomposition (7.73), that is,  $L = L' \oplus (L')^\perp$ . If  $L'$  is one-dimensional and spanned by a timelike vector or is a pseudo-Euclidean plane, then this is exactly decomposition (7.93) with  $L_0 = (L')^\perp$  and  $L_1 = L'$ . In the opposite case, the subspace  $L'$  is a Euclidean space of dimension 1 or 2, and the subspace  $(L')^\perp$  is a pseudo-Euclidean space of dimension  $n - 1$  or  $n - 2$  respectively. By the induction hypothesis, for  $(L')^\perp$ , we have the orthogonal decomposition  $(L')^\perp = L'_0 \oplus L'_1$  analogous to (7.93). From this, for  $L$  we obtain the decomposition (7.93) with  $L_0 = L' \oplus L'_0$  and  $L_1 = L'_1$ .

*Case 3.* The space  $L$  is cyclic for the transformation  $\mathcal{U}$ , corresponding to the unique real eigenvalue  $\lambda$  and principal vector  $e$  of grade  $m = n$ . Obviously, for  $n = 2$ , this is impossible: as we saw in Example 7.61, in a suitable basis of a pseudo-Euclidean plane, a Lorentz transformation has either diagonal form (7.87) or the form (7.88) with distinct eigenvalues  $\pm 1$ . In both cases, it is obvious that the pseudo-Euclidean plane  $L$  cannot be a cyclic subspace of the transformation  $\mathcal{U}$ .

Let us consider the case of a pseudo-Euclidean space  $L$  of dimension  $n \geq 3$ . We shall prove that  $L$  can be a cyclic subspace of the transformation  $\mathcal{U}$  only if  $n = 3$ .

As we established in Sect. 5.1, in a cyclic subspace  $L$ , there is a basis  $e_1, \dots, e_n$  defined by formula (5.5), that is,

$$e_1 = e, \quad e_2 = (\mathcal{U} - \lambda \mathcal{E})(e), \quad \dots, \quad e_n = (\mathcal{U} - \lambda \mathcal{E})^{n-1}(e), \quad (7.94)$$

in which relationships (5.6) hold:

$$\mathcal{U}(e_1) = \lambda e_1 + e_2, \quad \mathcal{U}(e_2) = \lambda e_2 + e_3, \quad \dots, \quad \mathcal{U}(e_n) = \lambda e_n. \quad (7.95)$$

In this basis, the matrix of the transformation  $\mathcal{U}$  has the form of a Jordan block

$$U = \begin{pmatrix} \lambda & 0 & 0 & \cdots & \cdots & 0 \\ 1 & \lambda & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \lambda & & & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \lambda & 0 \\ 0 & 0 & 0 & \cdots & 1 & \lambda \end{pmatrix}. \quad (7.96)$$

It is easy to see that the eigenvector  $\mathbf{e}_n$  is lightlike. Indeed, if we had  $(\mathbf{e}_n^2) \neq 0$ , then we would have the orthogonal decomposition  $L = \langle \mathbf{e}_n \rangle \oplus \langle \mathbf{e}_n \rangle^\perp$ , where both subspaces  $\langle \mathbf{e}_n \rangle$  and  $\langle \mathbf{e}_n \rangle^\perp$  are invariant. But this contradicts the assumption that the space  $L$  is cyclic.

Since  $\mathcal{U}$  is a Lorentz transformation, it preserves the inner product of vectors, and from (7.95), we obtain the equality

$$\begin{aligned} (\mathbf{e}_i, \mathbf{e}_n) &= (\mathcal{U}(\mathbf{e}_i), \mathcal{U}(\mathbf{e}_n)) = (\lambda \mathbf{e}_i + \mathbf{e}_{i+1}, \lambda \mathbf{e}_n) \\ &= \lambda^2 (\mathbf{e}_i, \mathbf{e}_n) + \lambda (\mathbf{e}_{i+1}, \mathbf{e}_n) \end{aligned} \quad (7.97)$$

for all  $i = 1, \dots, n-1$ .

If  $\lambda^2 \neq 1$ , then from (7.97), it follows that

$$(\mathbf{e}_i, \mathbf{e}_n) = \frac{\lambda}{1 - \lambda^2} (\mathbf{e}_{i+1}, \mathbf{e}_n).$$

Substituting into this equality the values of the index  $i = n-1, \dots, 1$ , taking into account that  $(\mathbf{e}_n^2) = 0$ , we therefore obtain step by step that  $(\mathbf{e}_i, \mathbf{e}_n) = 0$  for all  $i$ . This means that the eigenvector  $\mathbf{e}_n$  is contained in the radical of the space  $L$ , and since  $L$  is a pseudo-Euclidean space (that is, in particular, nondegenerate), it follows that  $\mathbf{e}_n = \mathbf{0}$ . This contradiction shows that  $\lambda^2 = 1$ .

Substituting  $\lambda^2 = 1$  into the equalities (7.97) and collecting like terms, we find that  $(\mathbf{e}_{i+1}, \mathbf{e}_n) = 0$  for all indices  $i = 1, \dots, n-1$ , that is,  $(\mathbf{e}_j, \mathbf{e}_n) = 0$  for all indices  $j = 2, \dots, n$ . In particular, we have the equalities  $(\mathbf{e}_{n-1}, \mathbf{e}_n) = 0$  for  $n > 2$  and  $(\mathbf{e}_{n-2}, \mathbf{e}_n) = 0$  for  $n > 3$ . From this it follows that  $n = 3$ . Indeed, from the condition of preservation of the inner product, we have the relationship

$$\begin{aligned} (\mathbf{e}_{n-2}, \mathbf{e}_{n-1}) &= (\mathcal{U}(\mathbf{e}_{n-2}), \mathcal{U}(\mathbf{e}_{n-1})) = (\lambda \mathbf{e}_{n-2} + \mathbf{e}_{n-1}, \lambda \mathbf{e}_{n-1} + \mathbf{e}_n) \\ &= \lambda^2 (\mathbf{e}_{n-2}, \mathbf{e}_{n-1}) + \lambda (\mathbf{e}_{n-2}, \mathbf{e}_n) + \lambda (\mathbf{e}_{n-1}^2) + (\mathbf{e}_{n-1}, \mathbf{e}_n), \end{aligned}$$

from which, taking into account the relationships  $\lambda^2 = 1$  and  $(\mathbf{e}_{n-1}, \mathbf{e}_n) = 0$ , we have the equality  $(\mathbf{e}_{n-2}, \mathbf{e}_n) + (\mathbf{e}_{n-1}^2) = 0$ . If  $n > 3$ , then  $(\mathbf{e}_{n-2}, \mathbf{e}_n) = 0$ , and from this, we obtain that  $(\mathbf{e}_{n-1}^2) = 0$ , that is, the vector  $\mathbf{e}_{n-1}$  is lightlike.

Let us examine the subspace  $L' = \langle \mathbf{e}_n, \mathbf{e}_{n-1} \rangle$ . It is obvious that it is invariant with respect to the transformation  $\mathcal{U}$ , and since it contains two linearly independent

lightlike vectors  $e_n$  and  $e_{n-1}$ , then by Lemma 7.53, the subspace  $L'$  is a pseudo-Euclidean space, and we obtain the decomposition  $L = L' \oplus (L')^\perp$  as a direct sum of two invariant subspaces. But this contradicts the fact that the space  $L$  is cyclic. Therefore, the transformation  $\mathcal{U}$  can have cyclic subspaces only of dimension 3.

Putting together cases 1, 2, and 3, and taking into account the induction hypothesis, we obtain the assertion of the theorem.  $\square$

Combining Theorems 7.27 and 7.65, we obtain the following corollary.

**Corollary 7.66** *For every transformation of a pseudo-Euclidean space, there exists an orthonormal basis in which the matrix of the transformation has block-diagonal form with blocks of the following types:*

1. blocks of order 1 with elements  $\pm 1$ ;
2. blocks of order 2 of type (7.29);
3. blocks of order 2 of type (7.91)–(7.92);
4. blocks of order 3 corresponding to a three-dimensional cyclic subspace with eigenvalue  $\pm 1$ .

*It follows from the law of inertia that the matrix of a Lorentz transformation can contain not more than one block of type 3 or 4.*

Let us note as well that a block of type 4 corresponding to a three-dimensional cyclic subspace cannot be brought into Jordan normal form in an orthonormal basis. Indeed, as we saw earlier, a block of type 4 is brought into Jordan normal form in the basis (7.94), where the eigenvector  $e_n$  is lightlike, and therefore, it cannot belong to any orthonormal basis.

With the proof of Theorem 7.65 we have established necessary conditions for a Lorentz transformation to have a cyclic subspace—in particular, its dimension must be 3, corresponding to an eigenvalue equal to  $\pm 1$ , and eigenvector that is lightlike. Clearly, these necessary conditions are not sufficient, since in deriving them, we used the equalities  $(e_i, e_k) = (\mathcal{U}(e_i), \mathcal{U}(e_k))$  for only some of the vectors of the basis (7.94). Let us show that Lorentz transformations with cyclic subspaces indeed exist.

**Example 7.67** Let us consider a vector space  $L$  of dimension  $n = 3$ . Let us choose in  $L$  a basis  $e_1, e_2, e_3$  and define a transformation  $\mathcal{U} : L \rightarrow L$  using relationships (7.95) with the number  $\lambda = \pm 1$ . Then the matrix of the transformation  $\mathcal{U}$  will take the form of a Jordan block with eigenvalue  $\lambda$ .

Let us choose the Gram matrix for a basis  $e_1, e_2, e_3$  such that  $L$  is given the structure of a pseudo-Euclidean space. With the proof of Theorem 7.65, we have found necessary conditions  $(e_2, e_3) = 0$  and  $(e_3^2) = 0$ . Let us set  $(e_1^2) = a$ ,  $(e_1, e_2) = b$ ,  $(e_1, e_3) = c$ , and  $(e_2^2) = d$ . Then the Gram matrix can be written as

$$A = \begin{pmatrix} a & b & c \\ b & d & 0 \\ c & 0 & 0 \end{pmatrix}. \quad (7.98)$$

On the other hand, as we know (see Example 7.51, p. 270), in  $L$  there exists an orthonormal basis in which the Gram matrix is diagonal and has determinant  $-1$ . Since the sign of the determinant of the Gram matrix is one and the same for all bases, it follows that  $|A| = -c^2d < 0$ , that is,  $c \neq 0$  and  $d > 0$ .

The conditions  $c \neq 0$  and  $d > 0$  are also sufficient for the vector space in which the inner product is given by the Gram matrix  $A$  in the form (7.98) to be a pseudo-Euclidean space. Indeed, choosing a basis  $\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3$  in which the quadratic form associated with the matrix  $A$  has canonical form (6.28), we see that the condition  $|A| < 0$  is satisfied by, besides a pseudo-Euclidean space, only a space in which  $(\mathbf{g}_i^2) = -1$  for all  $i = 1, 2, 3$ . But such a quadratic form is negative definite, that is,  $(\mathbf{x}^2) < 0$  for all vectors  $\mathbf{x} \neq \mathbf{0}$ , and this contradicts that  $(\mathbf{e}_2^2) = d > 0$ .

Let us now consider the equalities  $(\mathbf{e}_i, \mathbf{e}_k) = (\mathcal{U}(\mathbf{e}_i), \mathcal{U}(\mathbf{e}_k))$  for all indices  $i \leq k$  from 1 to 3. Taking into account  $\lambda^2 = 1$ ,  $(\mathbf{e}_2, \mathbf{e}_3) = 0$ , and  $(\mathbf{e}_3^2) = 0$ , we see that they are satisfied automatically except for the cases  $i = k = 1$  and  $i = 1, k = 2$ . These two cases give the relationships  $2\lambda b + d = 0$  and  $c + d = 0$ . Thus we may choose the number  $a$  arbitrarily, the number  $d$  to be any positive number, and set  $c = -d$  and  $b = -\lambda d/2$ . It is also not difficult to ascertain that linearly independent vectors  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$  satisfying such conditions in fact exist.

Just as in a Euclidean space, the presence of different orientations of a pseudo-Euclidean space determined by the value of  $\varepsilon(\mathcal{U})$  for the Lorentz transformation  $\mathcal{U}$  is connected with the concept of continuous deformation of a transformation (p. 230), which defines an equivalence relation on the set of transformations.

Let  $\mathcal{U}_t$  be a family of Lorentz transformations continuously depending on the parameter  $t$ . Then  $|\mathcal{U}_t|$  also depends continuously on  $t$ , and since the determinant of a Lorentz transformation is equal to  $\pm 1$ , the value of  $|\mathcal{U}_t|$  is constant for all  $t$ . Thus Lorentz transformations with determinants having opposite signs cannot be continuously deformed into each other. But in contrast to orthogonal transformations of a Euclidean space, Lorentz transformations  $\mathcal{U}_t$  have an additional characteristic, the number  $\nu(\mathcal{U}_t)$  (see the definition on p. 276). Let us show that like the determinant  $|\mathcal{U}_t|$ , the number  $\nu(\mathcal{U}_t)$  is also constant.

To this end, let us choose an arbitrary timelike vector  $\mathbf{e}$  and make use of Lemma 7.59. The vector  $\mathcal{U}_t(\mathbf{e})$  is also timelike, and moreover,  $\nu(\mathcal{U}_t) = +1$  if  $\mathbf{e}$  and  $\mathcal{U}_t(\mathbf{e})$  lie inside one pole of the light cone, that is,  $(\mathbf{e}, \mathcal{U}_t(\mathbf{e})) < 0$ , and  $\nu(\mathcal{U}_t) = -1$  if  $\mathbf{e}$  and  $\mathcal{U}_t(\mathbf{e})$  lie inside different poles, that is,  $(\mathbf{e}, \mathcal{U}_t(\mathbf{e})) > 0$ . It then remains to observe that the function  $(\mathbf{e}, \mathcal{U}_t(\mathbf{e}))$  depends continuously on the argument  $t$ , and therefore can change sign only if for some value of  $t$ , it assumes the value zero. But from inequality (7.82) for timelike vectors  $\mathbf{x} = \mathbf{e}$  and  $\mathbf{y} = \mathcal{U}_t(\mathbf{e})$ , there follows the inequality

$$(\mathbf{e}, \mathcal{U}_t(\mathbf{e}))^2 \geq (\mathbf{e}^2) \cdot (\mathcal{U}_t(\mathbf{e})^2) > 0,$$

showing that  $(\mathbf{e}, \mathcal{U}_t(\mathbf{e}))$  cannot be zero for any value of  $t$ .

Thus taking into account Theorem 7.63, we see that the number of equivalence classes of Lorentz transformations is certainly not less than four. Now we shall

show that there are exactly four. To begin with, we shall establish this for a pseudo-Euclidean plane, and thereafter shall prove it for a pseudo-Euclidean space of arbitrary dimension.

*Example 7.68* The matrices (7.91), (7.92) presenting all possible Lorentz transformations of a pseudo-Euclidean plane can be continuously deformed into the matrices

$$\begin{aligned} E &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, & F_1 &= \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \\ F_2 &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, & F_3 &= \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned} \quad (7.99)$$

respectively. Indeed, we obtain the necessary continuous deformation if in the matrices (7.91), (7.92) we replace the parameter  $\psi$  by  $(1-t)\psi$ , where  $t \in [0, 1]$ . It is also clear that none of the four matrices (7.99) can be continuously deformed into any of the others: any two of them differ either by the signs of their determinants or in that one of them preserves the poles of the light cone, while the other causes them to exchange places.

In the general case, we have an analogue of Theorem 7.28.

**Theorem 7.69** *Two Lorentz transformations  $\mathcal{U}_1$  and  $\mathcal{U}_2$  of a real pseudo-Euclidean space are continuously deformable into each other if and only if  $\varepsilon(\mathcal{U}_1) = \varepsilon(\mathcal{U}_2)$ .*

*Proof* Just as in the case of Theorem 7.28, we begin with a more specific assertion: we shall show that an arbitrary Lorentz transformation  $\mathcal{U}$  for which

$$\varepsilon(\mathcal{U}) = (|\mathcal{U}|, \nu(\mathcal{U})) = (+1, +1) \quad (7.100)$$

holds can be continuously deformed into  $\mathcal{E}$ . Invoking Theorem 7.65, let us examine the orthogonal decomposition (7.93), denoting by  $\mathcal{U}_i$  the restriction of the transformation  $\mathcal{U}$  to the invariant subspace  $L_i$ , where  $i = 0, 1$ . We shall investigate three cases in turn.

*Case 1.* In the decomposition (7.93), the dimension of the subspace  $L_1$  is equal to 1, that is,  $L_1 = \langle e \rangle$ , where  $(e^2) < 0$ . Then to the subspace  $L_1$ , there corresponds in the matrix of the transformation  $\mathcal{U}$  a block of order 1 with  $\sigma = +1$  or  $-1$ , and  $\mathcal{U}_0$  is an orthogonal transformation that depending on the sign of  $\sigma$ , can be proper or improper, so that the condition  $|\mathcal{U}| = \sigma |\mathcal{U}_0| = 1$  is satisfied. However, it is easy to see that for  $\sigma = -1$ , we have  $\nu(\mathcal{U}) = -1$  (since  $(e, \mathcal{U}(e)) > 0$ ), and therefore, the condition (7.100) leaves only the case  $\sigma = +1$ , and consequently, the orthogonal transformation  $\mathcal{U}_0$  is proper. Then  $\mathcal{U}_1$  is the identity transformation (of a one-dimensional space). By Theorem 7.28, an orthogonal transformation  $\mathcal{U}_0$  is

continuously deformable into the identity, and therefore, the transformation  $\mathcal{U}$  is continuously deformable into  $\mathcal{E}$ .

*Case 2.* In the decomposition (7.93), the dimension of the subspace  $L_1$  is equal to 2, that is,  $L_1$  is a pseudo-Euclidean plane. Then as we established in Examples 7.62 and 7.68, in some orthonormal basis of the plane  $L_1$ , the matrix of the transformation  $\mathcal{U}_1$  has the form (7.92) and is continuously deformable into one of the four matrices (7.99). It is obvious that the condition  $\nu(\mathcal{U}) = 1$  is associated with only the matrix  $E$  and one of the matrices  $F_2, F_3$ , namely the one in which the eigenvalues  $\pm 1$  correspond to the eigenvectors  $\mathbf{g}_\pm$  in such a way that  $(\mathbf{g}_+^2) < 0$  and  $(\mathbf{g}_-^2) > 0$ . In this case, it is obvious that we have the orthogonal decomposition  $L_1 = \langle \mathbf{g}_+ \rangle \oplus \langle \mathbf{g}_- \rangle$ .

If the matrix of the transformation  $\mathcal{U}_1$  is continuously deformable into  $E$ , then the orthogonal transformation  $\mathcal{U}_0$  is proper, and it follows that it is also continuously deformable into the identity, which proves our assertion.

If the matrix of the transformation  $\mathcal{U}_1$  is continuously deformable into  $F_2$  or  $F_3$ , then the orthogonal transformation  $\mathcal{U}_0$  is improper, and consequently, its matrix is continuously deformable into the matrix (7.32), which has the eigenvalue  $-1$  corresponding to some eigenvector  $\mathbf{h} \in L_0$ . From the orthogonal decomposition  $L = L_0 \oplus \langle \mathbf{g}_+ \rangle \oplus \langle \mathbf{g}_- \rangle$ , taking into account  $(\mathbf{g}_+^2) < 0$ , it follows that the invariant plane  $L' = \langle \mathbf{g}_-, \mathbf{h} \rangle$  is a Euclidean space. The matrix of the restriction of  $\mathcal{U}$  to  $L'$  is equal to  $-E$ , and is therefore continuously deformable into  $E$ . And this implies that the transformation  $\mathcal{U}$  is continuously deformable into  $\mathcal{E}$ .

*Case 3.* In the decomposition (7.93), the subspace  $L_1$  is a cyclic three-dimensional pseudo-Euclidean space with eigenvalue  $\lambda = \pm 1$ . This case was examined in detail in Example 7.67, and we will use the notation introduced there. It is obvious that the condition  $\nu(\mathcal{U}) = 1$  is satisfied only for  $\lambda = 1$ , since otherwise, the transformation  $\mathcal{U}_1$  takes the lightlike eigenvector  $\mathbf{e}_3$  to  $-\mathbf{e}_3$ , that is, it transposes the poles of the light cone. Thus condition (7.100) corresponds to the Lorentz transformation  $\mathcal{U}_1$  with the value  $\varepsilon(\mathcal{U}_1) = (+1, +1)$  and proper orthogonal transformation  $\mathcal{U}_0$ .

Let us show that such a transformation  $\mathcal{U}_1$  is continuously deformable into the identity. Since  $\mathcal{U}_0$  is obviously also continuously deformable into the identity, this will give us the required assertion.

Thus let  $\lambda = 1$ . We shall fix in  $L_1$  a basis  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$  satisfying the following conditions introduced in Example 7.67:

$$\begin{aligned} (\mathbf{e}_1^2) &= a, & (\mathbf{e}_1, \mathbf{e}_2) &= -\frac{d}{2}, \\ (\mathbf{e}_1, \mathbf{e}_3) &= -d, & (\mathbf{e}_2^2) &= d, & (\mathbf{e}_2, \mathbf{e}_3) &= (\mathbf{e}_3^2) = 0 \end{aligned} \tag{7.101}$$

with some numbers  $a$  and  $d > 0$ . The Gram matrix  $A$  in this basis has the form (7.98) with  $c = -d$  and  $b = -d/2$ , while the matrix  $U_1$  of the transformation  $\mathcal{U}_1$  has the form of a Jordan block.

Let  $\mathcal{U}_t$  be a linear transformation of the space  $L_1$  whose matrix in the basis  $e_1, e_2, e_3$  has the form

$$U_t = \begin{pmatrix} 1 & 0 & 0 \\ t & 1 & 0 \\ \varphi(t) & t & 1 \end{pmatrix}, \quad (7.102)$$

where  $t$  is a real parameter taking values from 0 to 1, and  $\varphi(t)$  is a continuous function of  $t$  that we shall choose in such a way that  $\mathcal{U}_t$  is a Lorentz transformation. As we know, for this, the relationship (7.85) with matrix  $U = U_t$  must be satisfied. Substituting in the equality  $U_t^* A U_t = A$  the matrix  $A$  of the form (7.98) with  $c = -d$  and  $b = -d/2$  and matrix  $U_t$  of the form (7.102) and equating corresponding elements on the left- and right-hand sides, we obtain that the equality  $U_t^* A U_t = A$  holds if  $\varphi(t) = t(t-1)/2$ . For such a choice of function  $\varphi(t)$ , we obtain a family of Lorentz transformations  $\mathcal{U}_t$  depending continuously on the parameter  $t \in [0, 1]$ . Moreover, it is obvious that for  $t = 1$ , the matrix  $U_t$  has the Jordan block  $U_1$ , while for  $t = 0$ , the matrix  $U_t$  equals  $E$ . Thus the family  $\mathcal{U}_t$  effects a continuous deformation of the transformation  $\mathcal{U}_1$  into  $\mathcal{E}$ .

Now let us prove the assertion of Theorem 7.69 in general form. Let  $\mathcal{W}$  be a Lorentz transformation with arbitrary  $\varepsilon(\mathcal{W})$ . We shall show that it can be continuously deformed into the transformation  $\mathcal{F}$ , having in some orthonormal basis the block-diagonal matrix

$$F = \begin{pmatrix} E & 0 \\ 0 & F' \end{pmatrix},$$

where  $E$  is the identity matrix of order  $n-2$  and  $F'$  is one of the four matrices (7.99). It is obvious that by choosing a suitable matrix  $F'$ , we may obtain the Lorentz transformation  $\mathcal{F}$  with any desired  $\varepsilon(\mathcal{F})$ . Let us select the matrix  $F'$  in such a way that  $\varepsilon(\mathcal{F}) = \varepsilon(\mathcal{W})$ .

Let us select in our space an arbitrary orthonormal basis, and in that basis, let the transformation  $\mathcal{W}$  have matrix  $W$ . Then the transformation  $\mathcal{U}$  having in this same basis the matrix  $U = WF$  is a Lorentz transformation, and moreover, by our choice of  $\varepsilon(\mathcal{F}) = \varepsilon(\mathcal{W})$ , we have the equality  $\varepsilon(\mathcal{U}) = \varepsilon(\mathcal{W})\varepsilon(\mathcal{F}) = (+1, +1)$ . Further, from the trivially verified relationship  $F^{-1} = F$ , we obtain  $W = UF$ , that is,  $\mathcal{W} = \mathcal{U}\mathcal{F}$ . We shall now make use of a family  $\mathcal{U}_t$  that effects the continuous deformation of the transformation  $\mathcal{U}$  into  $\mathcal{E}$ . From the equality  $\mathcal{W} = \mathcal{U}\mathcal{F}$ , with the help of Lemma 4.37, we obtain the relationship  $\mathcal{W}_t = \mathcal{U}_t\mathcal{F}$ , in which  $\mathcal{W}_0 = \mathcal{E}\mathcal{F} = \mathcal{F}$  and  $\mathcal{W}_1 = \mathcal{U}\mathcal{F} = \mathcal{W}$ . Thus it is this family  $\mathcal{W}_t = \mathcal{U}_t\mathcal{F}$  that accomplishes the deformation of the Lorentz transformation  $\mathcal{W}$  into  $\mathcal{F}$ .

If  $\mathcal{U}_1$  and  $\mathcal{U}_2$  are Lorentz transformations such that  $\varepsilon(\mathcal{U}_1) = \varepsilon(\mathcal{U}_2)$ , then by what we showed earlier, each of them is continuously deformable into  $\mathcal{F}$  with one and the same matrix  $F'$ . Consequently, by transitivity, the transformations  $\mathcal{U}_1$  and  $\mathcal{U}_2$  are continuously deformable into each other.  $\square$

Similarly to what we did in Sects. 4.4 and 7.3 for nonsingular and orthogonal transformations, we can express the fact established by Theorem 7.69 in topological

form: the set of Lorentz transformations of a pseudo-Euclidean space of a given dimension has *exactly four* path-connected components. They correspond to the four possible values of  $\varepsilon(\mathcal{U})$ .

Let us note that the existence of four (instead of two) orientations is not a specific property of pseudo-Euclidean spaces with the quadratic form (7.76), as was the case with the majority of properties of this section. It holds for all vector spaces with a bilinear inner product  $(\mathbf{x}, \mathbf{y})$ , provided that it is nonsingular and the quadratic form  $(\mathbf{x}^2)$  is neither positive nor negative definite. We can indicate (without pretending to provide a proof) the reason for this phenomenon. If the form  $(\mathbf{x}^2)$ , in canonical form, appears as

$$x_1^2 + \cdots + x_s^2 - x_{s+1}^2 - \cdots - x_n^2, \quad \text{where } s \in \{1, \dots, n-1\},$$

then the transformations that preserve it include first of all, the orthogonal transformations preserving the form  $x_1^2 + \cdots + x_s^2$  and not changing the coordinates  $x_{s+1}, \dots, x_n$ , and secondly, the transformations preserving the quadratic form  $x_{s+1}^2 + \cdots + x_n^2$  and not changing the coordinates  $x_1, \dots, x_s$ . Every type of transformation is “responsible” for its own orientation.



## Chapter 8

# Affine Spaces

The usual objects of study in plane and solid geometry are the plane and three-dimensional space, both of which consist of *points*. However, *vector* spaces are logically simpler, and therefore, we began by studying them. Now we can move on to “point” (affine) spaces. The theory of such spaces is closely related to that of vector spaces, and so in this chapter, we shall be concerned only with questions relating specifically to this case.

### 8.1 The Definition of an Affine Space

Let us return to the starting point in the theory of vector spaces, namely to Sect. 3.1. There, we said that two points in the plane (or in space) determine a vector. We shall make this property the basis of the axiomatic definition of affine spaces.

**Definition 8.1** An *affine space* is a pair  $(V, L)$  consisting of a set  $V$  (whose elements are called *points*) and a vector space  $L$ , on which a rule is defined whereby two points  $A, B \in V$  are associated with a vector of the space  $L$ , which we shall denote by  $\overrightarrow{AB}$  (the order of the points  $A$  and  $B$  is significant). Here the following conditions must be satisfied:

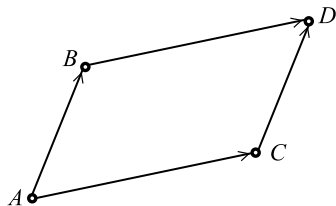
- (1)  $\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}$ .
- (2) For every three points  $A, B, C \in V$ , there exists a unique point  $D \in V$  such that

$$\overrightarrow{AB} = \overrightarrow{CD}. \quad (8.1)$$

- (3) For every two points  $A, B \in V$  and scalar  $\alpha$ , there exists a unique point  $C \in V$  such that

$$\overrightarrow{AC} = \alpha \overrightarrow{AB}. \quad (8.2)$$

**Remark 8.2** From condition (2), it follows that we also have  $\overrightarrow{AC} = \overrightarrow{BD}$ . Indeed, in view of condition (1), we have the equalities  $\overrightarrow{AB} + \overrightarrow{BD} = \overrightarrow{AD}$  and  $\overrightarrow{AC} + \overrightarrow{CD} =$

**Fig. 8.1** Equality of vectors

$\overrightarrow{AD}$ . This implies that  $\overrightarrow{AB} + \overrightarrow{BD} = \overrightarrow{AC} + \overrightarrow{CD}$  (see Fig. 8.1). Since  $\overrightarrow{AB} = \overrightarrow{CD}$  by assumption, and all vectors belong to the space  $L$ , it follows that  $\overrightarrow{AC} = \overrightarrow{BD}$ .

From these conditions and the definition of a vector space, it is easy to derive that for an arbitrary point  $A \in V$ , the vector  $\overrightarrow{AA}$  is equal to  $\mathbf{0}$ , and for every pair of points  $A, B \in V$ , we have the equality

$$\overrightarrow{BA} = -\overrightarrow{AB}.$$

It is equally easy to verify that if we are given a point  $A \in V$  and a vector  $\mathbf{x} = \overrightarrow{AB}$  in the space  $L$ , then the point  $B \in V$  is thereby uniquely determined.

**Theorem 8.3** *The totality of all vectors of the form  $\overrightarrow{AB}$ , where  $A$  and  $B$  are arbitrary points of  $V$ , forms a subspace  $L'$  of the space  $L$ .*

*Proof* Let  $\mathbf{x} = \overrightarrow{AB}$ ,  $\mathbf{y} = \overrightarrow{CD}$ . By condition (2), there exists a point  $K$  such that  $\overrightarrow{BK} = \overrightarrow{CD}$ . Then by condition (1), the vector

$$\overrightarrow{AK} = \overrightarrow{AB} + \overrightarrow{BK} = \overrightarrow{AB} + \overrightarrow{CD} = \mathbf{x} + \mathbf{y}$$

is again contained in the subspace  $L'$ . Analogously, for any vector  $\mathbf{x} = \overrightarrow{AB}$  in  $L'$ , condition (3) gives the vector  $\overrightarrow{AC} = \alpha \overrightarrow{AB} = \alpha \mathbf{x}$ , which consequently also is contained in  $L'$ .  $\square$

In view of Theorem 8.3, we shall require for the study of an affine space  $(V, L)$  not all the vectors of the space  $L$ , but only those that lie in the subspace  $L'$ . Therefore, in what follows, we shall denote the space  $L'$  by  $L$ . In other words, we shall assume that the following condition is satisfied: for every vector  $\mathbf{x} \in L$ , there exist points  $A$  and  $B$  in  $V$  such that  $\mathbf{x} = \overrightarrow{AB}$ .

This condition does not impose any additional constraints. It is simply equivalent to a change of notation:  $L$  instead of  $L'$ .

**Example 8.4** Every vector space  $L$  defines an affine space  $(L, L)$  if for two vectors  $\mathbf{a}, \mathbf{b} \in L$  considered as points of the set  $V = L$ , we set  $\overrightarrow{\mathbf{a}\mathbf{b}} = \mathbf{b} - \mathbf{a}$ . In particular, the totality  $\mathbb{K}^n$  of all rows of length  $n$  defines an affine space.

**Example 8.5** The *plane* and *space* studied in a course in elementary or analytic geometry are examples of affine spaces.

Condition (2) in the definition of an affine space shows that no matter how we choose the point  $O$  in the set  $V$ , every vector  $\mathbf{x} \in L$  can be represented as  $\mathbf{x} = \overrightarrow{OA}$ . Moreover, from the requirement of the uniqueness of the point  $D$  in condition (2), it follows that for a designated point  $O$  and vector  $\mathbf{x}$ , the point  $A$  is uniquely determined by the condition  $\overrightarrow{OA} = \mathbf{x}$ . Thus having chosen (arbitrarily) a point  $O \in V$  and associating with each point  $A \in V$  the vector  $\overrightarrow{OA}$ , we obtain a bijection between the points  $A$  of the set  $V$  and the vectors  $\mathbf{x}$  of the space  $L$ . In other words, an affine space is a vector space in which the coordinate origin  $O$  is not fixed. This notion is more natural from a physical point of view; in an affine space, all points are created equal, or in other words, the space is *uniform*. Mathematically, such a notion seems more complex: we need to specify not one, but two sets:  $V$  and  $L$ . And though we write an affine space as a pair  $(V, L)$ , we shall often denote such a space simply by  $V$ , leaving  $L$  implied and assuming that the condition formulated above is satisfied. In this case, we shall call  $L$  the *space of vectors* of the affine space  $V$ .

**Definition 8.6** The *dimension* of an affine space  $(V, L)$  is the dimension of the associated vector space  $L$ . When we wish to focus our attention on the space  $V$ , then we shall denote the dimension by  $\dim V$ .

In the sequel, we shall consider only spaces of finite dimension. We shall call an affine space of dimension 1 a *line*, and an affine space of dimension 2, a *plane*.

Having selected the point  $O \in V$ , we obtain a bijection  $V \rightarrow L$ . If  $\dim L = n$  and we choose in the space  $L$  some basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , then we have the isomorphism  $L \cong \mathbb{K}^n$ . Thus for an arbitrary choice of a point  $O \in V$  and basis in  $L$ , we obtain a bijection  $V \rightarrow \mathbb{K}^n$  and define each point of the affine space  $V$  by the set of coordinates  $(\alpha_1, \dots, \alpha_n)$  of the vector  $\mathbf{x} = \overrightarrow{OA}$  in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ .

**Definition 8.7** The point  $O$  and basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  together are called a *frame of reference* in the space  $V$ , and we write  $(O; \mathbf{e}_1, \dots, \mathbf{e}_n)$ . The  $n$ -tuple  $(\alpha_1, \dots, \alpha_n)$  associated with the point  $A \in V$  is called the *coordinates* of the point  $A$  of the associated frame of reference.

If relative to the frame of reference  $(O; \mathbf{e}_1, \dots, \mathbf{e}_n)$ , the point  $A$  has coordinates  $(\alpha_1, \dots, \alpha_n)$ , while the point  $B$  has coordinates  $(\beta_1, \dots, \beta_n)$ , then the vector  $\overrightarrow{AB}$  has, with respect to the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , coordinates  $(\beta_1 - \alpha_1, \dots, \beta_n - \alpha_n)$ .

Just as with the selection of a basis in a vector space, every vector of that space is determined by its coordinates, likewise is every point of an affine space determined by its coordinates in a given frame of reference. Thus a frame of reference plays the same role in the theory of affine spaces as that played by a basis in the theory of vector spaces. We have defined frame of reference as a collection consisting of the point  $O$  and  $n$  vectors  $\mathbf{e}_1, \dots, \mathbf{e}_n$  that form a basis of  $L$ . Any of these vectors  $\mathbf{e}_i$  can be written in the form  $\mathbf{e}_i = \overrightarrow{OA_i}$ , and then it is possible to give the frame of reference

as a collection of  $n + 1$  points  $O, A_1, \dots, A_n$ . Here the points  $O, A_1, \dots, A_n$  are not arbitrary; they must satisfy the property that the vectors  $\overrightarrow{OA_1}, \dots, \overrightarrow{OA_n}$  form a basis of  $L$ , that is, they must be linearly independent.

We have seen that the choice of a point  $O$  in  $V$  determines an isomorphism between  $V$  and  $L$  that assigns to each point  $A \in V$  the vector  $\overrightarrow{OA} \in L$ . Let us consider how this correspondence changes when we change the point  $O$ . If we began with the point  $O'$ , then we will have placed in correspondence with the point  $A$ , the vector  $\overrightarrow{O'A}$ , which, by definition of an affine space, is equal to  $\overrightarrow{O'O} + \overrightarrow{OA}$ . Thus if in the first case, we assign to the point  $A$  the vector  $\mathbf{x}$ , then in the second, we assign the vector  $\mathbf{x} + \mathbf{a}$ , where  $\mathbf{a} = \overrightarrow{O'O}$ . We obtain a corresponding mapping of the set  $V$  if to the point  $A$ , we assign the point  $B$  such that  $\overrightarrow{AB} = \mathbf{a}$ . Such a point  $B$  is uniquely determined by the choice of  $A$  and  $\mathbf{a}$ .

**Definition 8.8** A translation of an affine space  $(V, L)$  by a vector  $\mathbf{a} \in L$  is a mapping of the set  $V$  into itself that assigns to the point  $A$  the point  $B$  such that  $\overrightarrow{AB} = \mathbf{a}$ . (The existence and uniqueness of such a point  $B \in V$  for every  $A \in V$  and  $\mathbf{a} \in L$  follows from the definition of affine space.)

We shall denote the translation by the vector  $\mathbf{a}$  by  $\mathcal{T}_a$ . Thus the definition of a translation can be written as the formula

$$\mathcal{T}_a(A) = B, \quad \text{where } \overrightarrow{AB} = \mathbf{a}.$$

From the given definition, a translation is an isomorphism of the set  $V$  into itself. It can be depicted with the help of the diagram

$$\begin{array}{ccc} V & & \\ \mathcal{T}_a \downarrow & \searrow \psi & \\ & L & \\ & \nearrow \psi' & \\ V & & \end{array} \quad (8.3)$$

where the bijection  $\psi$  between  $V$  and  $L$  is defined using the point  $O$ , while the bijection  $\psi'$  uses the point  $O'$ , and  $\mathcal{T}_a$  is a translation by the vector  $\mathbf{a} = \overrightarrow{O'O}$ . As a result, the mapping  $\psi$  is the product (sequential application, or composition) of the mappings  $\mathcal{T}_a$  and  $\psi'$ . This relationship can be more briefly written as  $\psi' = \psi + \mathbf{a}$ .

**Proposition 8.9** Translations possess the following properties:

- (1)  $\mathcal{T}_a \mathcal{T}_b = \mathcal{T}_{a+b}$ ,
- (2)  $\mathcal{T}_0 = \mathcal{E}$ ,
- (3)  $\mathcal{T}_{-\mathbf{a}} = \mathcal{T}_a^{-1}$ .

*Proof* In property (1), the left-hand side consists of the product of mappings, which means that for every point  $C \in V$ , the equality

$$\mathcal{T}_a(\mathcal{T}_b(C)) = \mathcal{T}_{a+b}(C) \quad (8.4)$$

is satisfied. Let us represent the vector  $\mathbf{b}$  in the form  $\mathbf{b} = \overrightarrow{CP}$  (not only is this possible, but by the definition of affine space, the point  $P \in V$  is uniquely determined). Then we have the equality  $\mathcal{T}_b(C) = P$ . Likewise, let us represent the vector  $\mathbf{a}$  in the form  $\mathbf{a} = \overrightarrow{PQ}$ . Then analogously,  $\mathcal{T}_a(P) = Q$ . It follows from these relationships that

$$\mathbf{a} + \mathbf{b} = \overrightarrow{CP} + \overrightarrow{PQ} = \overrightarrow{CQ},$$

from which we obviously obtain  $\mathcal{T}_{a+b}(C) = Q$ . On the other hand, we have the equality  $\mathcal{T}_a(\mathcal{T}_b(C)) = \mathcal{T}_a(P) = Q$ , which proves the relationship (8.4).

Properties (2) and (3) can be proved even more easily.  $\square$

Let us note that for any two points  $A, B \in V$ , there exists a unique vector  $\mathbf{a} \in L$  for which  $\mathcal{T}_a(A) = B$ , namely, the vector  $\mathbf{a} = \overrightarrow{AB}$ .

Suppose that we are given a certain frame of reference  $(O; \mathbf{e}_1, \dots, \mathbf{e}_n)$ . Relative to this frame of reference, every point  $A \in V$  has coordinates  $(x_1, \dots, x_n)$ . A function  $F(A)$  defined on the affine space  $V$  and taking numeric values is called a *polynomial* if it can be written as a polynomial in the coordinates  $x_1, \dots, x_n$ .

This definition can be given a different formulation. Let us denote by  $\psi: V \rightarrow L$  the bijection between  $V$  and  $L$  determined by the selection of an arbitrary point  $O$ . Then the function  $F$  on  $V$  is a polynomial if it can be represented in the form  $F(A) = G(\psi(A))$ , where  $G(\mathbf{x})$  is a polynomial on the space  $L$  (see the definition on p. 127). To be sure, it is still necessary to verify that this definition does not depend on the choice of frame of reference  $(O; \mathbf{e}_1, \dots, \mathbf{e}_n)$ , but this can be done very easily. If  $\psi': V \rightarrow L$  is a bijection between  $V$  and  $L$  determined by the choice of point  $O'$  (cf. diagram (8.3)), then  $\psi' = \psi + \mathbf{a}$ . As we saw in Sect. 3.8, the property of a function  $G(\mathbf{x})$  being a polynomial does not depend on the choice of basis in  $L$ , and it remains to verify that for a polynomial  $G(\mathbf{x})$  and vector  $\mathbf{a} \in L$ , the function  $G(\mathbf{x} + \mathbf{a})$  is also a polynomial. It is clearly sufficient to verify this for the monomial  $c x_1^{k_1} \dots x_n^{k_n}$ . If the vector  $\mathbf{x}$  has coordinates  $x_1, \dots, x_n$ , and the vector  $\mathbf{a}$  has coordinates  $a_1, \dots, a_n$ , then substituting them into the monomial  $c x_1^{k_1} \dots x_n^{k_n}$ , we obtain the expression  $c(x_1 + a_1)^{k_1} \dots (x_n + a_n)^{k_n}$ , which is clearly also a polynomial in the variables  $x_1, \dots, x_n$ .

Using the same considerations as those employed in Example 3.86 on p. 130, we may define for an arbitrary polynomial  $F$  on an affine space  $V$  its *differential*  $d_O F$  at an arbitrary point  $O \in V$ . Here the differential  $d_O F$  will be a linear function on the space of vectors  $L$  of the space  $V$ , that is, it will be a vector in the dual space  $L^*$ . Indeed, let us consider the bijection  $\psi: V \rightarrow L$  constructed earlier, for which  $\psi(O) = \mathbf{0}$ ; let us represent  $F$  in the form  $F(A) = G(\psi(A))$ , where  $G(\mathbf{x})$  is some polynomial on the vector space  $L$ ; and let us define  $d_O F = d_{\mathbf{0}} G$  as a linear function on  $L$ .

Suppose that we are given the frame of reference  $(O; \mathbf{e}_1, \dots, \mathbf{e}_n)$  in the space  $V$ . Then  $F(A)$  is a polynomial in the coordinates of the point  $A$  with respect to this frame of reference. Let us write down the expression  $d_O F$  in these coordinates. By definition, the differential

$$d_O F = d_0 G = \sum_{i=1}^n \frac{\partial G}{\partial x_i}(\mathbf{0})x_i$$

is a linear function in the coordinates  $x_1, \dots, x_n$  with respect to the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ . Here  $\partial G / \partial x_i$  is a polynomial, and it corresponds to some polynomial  $\Phi_i$  on  $V$ , that is, it has the form  $\Phi_i(A) = \frac{\partial G}{\partial x_i}(\psi(A))$ . By definition, we set  $\Phi_i = \partial F / \partial x_i$ . It is easy to verify that if we express  $F$  and  $\Phi_i$  as polynomials in  $x_1, \dots, x_n$ , then  $\Phi_i$  will indeed be the partial derivative of  $F$  with respect to the variable  $x_i$ . Since  $\psi(O) = \mathbf{0}$ , it follows that  $\frac{\partial G}{\partial x_i}(\mathbf{0}) = \frac{\partial F}{\partial x_i}(O)$ . Consequently, we obtain for the differential  $d_O F$ , the expression

$$d_O F = \sum_{i=1}^n \frac{\partial F}{\partial x_i}(O)x_i,$$

which is similar to formula (3.70) obtained in Sect. 3.8.

## 8.2 Affine Spaces

**Definition 8.10** A subset  $V' \subset V$  of an affine space  $(V, L)$  is an *affine subspace* if the set of vectors  $\overrightarrow{AB}$  for all  $A, B \in V'$  forms a vector subspace  $L'$  of the vector space  $L$ .

It is obvious that then  $V'$  itself is an affine subspace, and  $L'$  is its space of vectors. If  $\dim V' = \dim V - 1$ , then  $V'$  is called a *hyperplane* in  $V$ .

*Example 8.11* A typical example of an affine subspace is the set  $V'$  of solutions of the system of linear equations (1.3). If the coefficients  $a_{ij}$  and constants  $b_i$  of the system of equations (1.3) lie in the field  $\mathbb{K}$ , then the set of solutions  $V'$  is contained in the set of rows  $\mathbb{K}^n$  of length  $n$ , which we view as an affine space  $(\mathbb{K}^n, \mathbb{K}^n)$ , that is,  $V = \mathbb{K}^n$  and  $L = \mathbb{K}^n$ .

For a proof of the fact that the solution set  $V'$  is an affine subspace, let us verify that its space of vectors  $L'$  is the solution space of the homogeneous system of linear equations associated with (1.3). That the set of solutions of a linear homogeneous system is a vector subspace of  $\mathbb{K}^n$  was established in Sect. 3.1 (Example 3.8). Let the rows  $\mathbf{x}$  and  $\mathbf{y}$  be solutions of the system (1.3), viewed now as points of the affine space  $V = \mathbb{K}^n$ . We must verify that the vector  $\overrightarrow{\mathbf{x}\mathbf{y}}$  defined as in the above example is contained in  $L'$ . But in accordance with this example, we must set  $\overrightarrow{\mathbf{x}\mathbf{y}} = \mathbf{y} - \mathbf{x}$ , and it then remains for us to verify that the row  $\mathbf{y} - \mathbf{x}$  belongs to the subspace  $L'$ , that is, it is a solution of the homogeneous system associated with the system (1.3).

It suffices to verify this property separately for each equation. Let the  $i$ th equation of the linear homogeneous system associated with (1.3) be given in the form (1.10), that is,  $F_i(\mathbf{x}) = 0$ , where  $F_i$  is some linear function. By assumption,  $\mathbf{x}$  and  $\mathbf{y}$  are solutions of the system (1.3), in particular,  $F_i(\mathbf{x}) = b_i$  and  $F_i(\mathbf{y}) = b_i$ . From this it follows that  $F_i(\mathbf{y} - \mathbf{x}) = F_i(\mathbf{y}) - F_i(\mathbf{x}) = b_i - b_i = 0$ , as asserted.

**Example 8.12** Let us now prove that conversely, every affine subspace of the affine space  $(\mathbb{K}^n, \mathbb{K}^n)$  is defined by linear equations, that is, if  $V'$  is an affine subspace, then  $V'$  coincides with the set of solutions of some system of linear equations. Since  $V'$  is a subspace of the affine space  $(\mathbb{K}^n, \mathbb{K}^n)$ , it follows by definition that the corresponding set of vectors  $L'$  is a subspace of the vector space  $\mathbb{K}^n$ . We saw in Sect. 3.1 (Example 3.8) that it is then defined in  $\mathbb{K}^n$  by a homogeneous system of linear equations

$$F_1(\mathbf{x}) = 0, \quad F_2(\mathbf{x}) = 0, \quad \dots, \quad F_m(\mathbf{x}) = 0. \quad (8.5)$$

Let us consider an arbitrary point  $A \in V'$  and set  $F_i(A) = b_i$  for all  $i = 1, \dots, m$ . We shall prove that then the subspace  $V'$  coincides with the set of solutions of the system

$$F_1(\mathbf{x}) = b_1, \quad F_2(\mathbf{x}) = b_2, \quad \dots, \quad F_m(\mathbf{x}) = b_m. \quad (8.6)$$

Indeed, let us take an arbitrary point  $B \in V'$ . Let the points  $A$  and  $B$  have coordinates  $A = (\alpha_1, \dots, \alpha_n)$  and  $B = (\beta_1, \dots, \beta_n)$  in some frame of reference. Then the coordinates of the vector  $\overrightarrow{AB}$  are equal to  $(\beta_1 - \alpha_1, \dots, \beta_n - \alpha_n)$ , and we know that the point  $B$  belongs to  $V'$  if and only if the vector  $\mathbf{x} = \overrightarrow{AB}$  belongs to the subspace  $L'$ , that is, satisfies equations (8.5). Now using the fact that the functions  $F_i$  are linear, we obtain that for any one of them,

$$F_i(\beta_1 - \alpha_1, \dots, \beta_n - \alpha_n) = F_i(\beta_1, \dots, \beta_n) - F_i(\alpha_1, \dots, \alpha_n) = F_i(B) - b_i.$$

This implies that the point  $B$  belongs to the affine subspace  $V'$  if and only if  $F_i(B) = b_i$ , that is, its coordinates satisfy equations (8.6).

**Definition 8.13** Affine subspaces  $V'$  and  $V''$  are said to be *parallel* if they have the same set of vectors, that is, if  $L' = L''$ .

It is easy to see that two parallel subspaces either have no points in common or else coincide. Indeed, suppose that  $V'$  and  $V''$  are parallel and the point  $A$  belongs to  $V' \cap V''$ . Since the spaces of vectors for  $V'$  and  $V''$  coincide, it follows that for an arbitrary point  $B \in V'$ , there exists a point  $C \in V''$  such that  $\overrightarrow{AB} = \overrightarrow{AC}$ . Hence, taking into account the uniqueness of the point  $D$  in the relationship (8.1) from the definition of an affine space, it follows that  $B = C$ , which implies that  $V' \subset V''$ . Since the definition of parallelism does not depend on the order of the subspaces  $V'$  and  $V''$ , the opposite inclusion  $V'' \subset V'$  holds as well, which yields that  $V' = V''$ .

Let  $V'$  and  $V''$  be two parallel subspaces, and let us choose in each of them a point:  $A \in V'$  and  $B \in V''$ . Setting the vector  $\overrightarrow{AB}$  equal to  $\mathbf{a}$ , we obtain, by definition of the translation  $\mathcal{T}_{\mathbf{a}}$ , that  $\mathcal{T}_{\mathbf{a}}(A) = B$ .

Let us consider an arbitrary point  $C \in V'$ . It follows from the definition of parallelism that there exists a point  $D \in V''$  such that  $\overrightarrow{AC} = \overrightarrow{BD}$ . From this, it follows easily that  $\overrightarrow{CD} = \overrightarrow{AB} = \mathbf{a}$ ; see Fig. 8.1 and Remark 8.2. But this implies that  $\mathcal{T}_{\mathbf{a}}(C) = D$ . In other words,  $\mathcal{T}_{\mathbf{a}}(V') \subset V''$ . Similarly, we obtain that  $\mathcal{T}_{-\mathbf{a}}(V'') \subset V'$ , whence from properties 1, 2, and 3 of a translation, it follows that  $V'' \subset \mathcal{T}_{\mathbf{a}}(V')$ . This implies that  $\mathcal{T}_{\mathbf{a}}(V') = V''$ , that is, any two parallel subspaces can be mapped into each other by a translation. Conversely, it is easy to verify that affine subspaces  $V'$  and  $\mathcal{T}_{\mathbf{a}}(V')$  are parallel for any choice of  $V'$  and  $\mathbf{a}$ .

Let us consider two different points  $A$  and  $B$  of an affine space  $(V, L)$ . Then the totality of all points  $C$  whose existence is established by condition (3) in the definition of affine space (with arbitrary scalars  $\alpha$ ) forms, as is easy to see, an affine subspace  $V'$ . The corresponding vector subspace  $L'$  coincides with  $\langle \overrightarrow{AB} \rangle$ . Therefore,  $L'$ , and hence also the affine space  $(V', L')$ , is one-dimensional. It is called the *line passing through the points  $A$  and  $B$* .

The notion of a line is related to the general notion of affine subspace by the following result.

**Theorem 8.14** *In order for a subset  $M$  of an affine space  $V$  defined over a field of characteristic different from 2 to be an affine subspace of  $V$ , it is necessary and sufficient that for every two points of  $M$ , the line passing through them be entirely contained in  $M$ .*

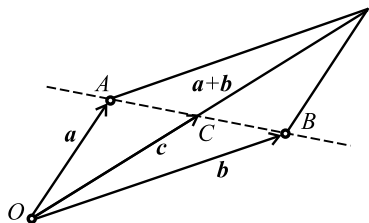
*Proof* The necessity of this condition is obvious. Let us prove its sufficiency. Let us choose an arbitrary point  $O \in M$ . We need to prove that the set of vectors  $\overrightarrow{OA}$ , where  $A$  runs over all possible points of the set  $M$ , forms a subspace  $L'$  of the space of vectors  $L$  of the affine space  $(V, L)$ . Then for any other point  $B \in M$ , the vector  $\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA}$  will lie in the subspace  $L'$ , whence  $(M, L')$  will be an affine subspace of the space  $(V, L)$ .

That the product of an arbitrary vector  $\overrightarrow{OA}$  and arbitrary scalar  $\alpha$  lies in  $L'$  derives from the condition that the line  $\langle \overrightarrow{OA} \rangle$  is contained in  $L'$ . Let us verify that the sum of two vectors  $\mathbf{a} = \overrightarrow{OA}$  and  $\mathbf{b} = \overrightarrow{OB}$  contained in  $L'$  is also contained in  $L'$ . For this, we shall need the condition that we required on the set of points of a line only for  $\alpha = 1/2$  (in order for us to be able to apply this condition, we have assumed that the field  $\mathbb{K}$  over which the affine space  $V$  in question is defined is of characteristic different from 2). Let  $C$  be a point of the line passing through  $A$  and  $B$  such that  $\overrightarrow{AC} = \frac{1}{2}\overrightarrow{AB}$ . By definition, along with each pair of points  $A$  and  $B$  of the set  $M$ , the line passing through them also belongs to this set. Hence it follows in particular that we have  $C \in M$  and  $\overrightarrow{OC} \in L'$ . Let us denote the vector  $\overrightarrow{OC}$  by  $\mathbf{c}$ ; see Fig. 8.2. Then we have the equalities

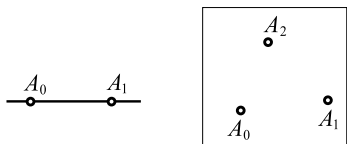
$$\mathbf{b} = \overrightarrow{OB} = \overrightarrow{OA} + \overrightarrow{AB} = \mathbf{a} + \overrightarrow{AB}, \quad \mathbf{c} = \overrightarrow{OC} = \overrightarrow{OA} + \overrightarrow{AC} = \mathbf{a} + \overrightarrow{AC},$$



**Fig. 8.2** Vectors  $\overrightarrow{OA}$ ,  $\overrightarrow{OB}$ , and  $\overrightarrow{OC}$



**Fig. 8.3** Independent points



and thus in our case, we have  $\overrightarrow{AB} = \mathbf{b} - \mathbf{a}$  and  $\overrightarrow{AC} = \mathbf{c} - \mathbf{a}$ , which implies  $\mathbf{c} - \mathbf{a} = \frac{1}{2}(\mathbf{b} - \mathbf{a})$ , that is,  $\mathbf{c} = \frac{1}{2}(\mathbf{a} + \mathbf{b})$ . Consequently, the vector  $\mathbf{a} + \mathbf{b}$  equals  $2\mathbf{c}$ , and since  $\mathbf{c}$  is in  $L'$ , the vector  $\mathbf{a} + \mathbf{b}$  is also in  $L'$ .  $\square$

Now let  $A_0, A_1, \dots, A_m$  be a collection of  $m + 1$  points in the affine space  $V$ . Let us consider the subspace

$$L' = \langle \overrightarrow{A_0A_1}, \overrightarrow{A_0A_2}, \dots, \overrightarrow{A_0A_m} \rangle$$

of the space  $L$ . It does not depend on the choice of point  $A_0$  among the given points  $A_0, A_1, \dots, A_m$ , and we may write it, for example, in the form  $\langle \dots, \overrightarrow{A_iA_j}, \dots \rangle$  for all  $i$  and  $j$ ,  $0 \leq i, j \leq m$ . The set  $V'$  of all points  $B \in V$  for which the vector  $\overrightarrow{A_0B}$  is in  $L'$  forms an affine subspace whose space of vectors is  $L'$ . By definition,  $\dim V' \leq m$ , and moreover,  $\dim V' = m$  if and only if  $\dim L' = m$ , that is, the vectors  $\overrightarrow{A_0A_1}, \overrightarrow{A_0A_2}, \dots, \overrightarrow{A_0A_m}$  are linearly independent. This provides the basis for the following definition.

**Definition 8.15** Points  $A_0, A_1, \dots, A_m$  of an affine space  $V$  for which

$$\dim \langle \overrightarrow{A_0A_1}, \overrightarrow{A_0A_2}, \dots, \overrightarrow{A_0A_m} \rangle = m$$

are called *independent*.

For example, the points  $A_0, A_1, \dots, A_n$  (where  $n = \dim V$ ) determine a frame of reference if and only if they are independent. Two distinct points are independent, as are three noncollinear points, and so on. See Fig. 8.3.

The following theorem gives an important property of affine spaces, connecting them with the familiar space of elementary geometry.

**Theorem 8.16** *There is a unique line passing through every pair of distinct points  $A$  and  $B$  of an affine space  $V$ .*

*Proof* It is obvious that distinct points  $A$  and  $B$  are independent, and the line  $V' \subset V$  containing them must coincide with the set of points  $C \in V$  for which  $\overrightarrow{AC} \in \langle \overrightarrow{AB} \rangle$  (instead of  $\overrightarrow{AC}$ , one could consider the vector  $\overrightarrow{BC}$ ; it determines the same subspace  $V' \subset V$ ). If  $\overrightarrow{AC} = \alpha \overrightarrow{AB}$  and  $\overrightarrow{AC'} = \beta \overrightarrow{AB}$ , then  $\overrightarrow{CC'} = (\beta - \alpha) \overrightarrow{AB}$ , whence it follows that  $V'$  is a line.  $\square$

Having selected on any line  $P$  of the affine space  $V$  the point  $O$  (reference point) and arbitrary point  $E \in P$  not equal to  $O$  (scale of measurement), we obtain for an arbitrary point  $A \in P$  the relationship

$$\overrightarrow{OA} = \alpha \overrightarrow{OE}, \quad (8.7)$$

where  $\alpha$  is some scalar, that is, an element of the field  $\mathbb{K}$  over which the affine space  $V$  under consideration is defined. The assignment  $A \mapsto \alpha$ , as is easily verified, establishes a bijection between the points  $A \in P$  and scalars  $\alpha$ . This correspondence, of course, depends on the choice of points  $O$  and  $E$  on the line. In fact, we have here a special case of the notion of coordinates relative to a frame of reference  $(O; \mathbf{e})$  on the affine line  $P$ , where  $\mathbf{e} = \overrightarrow{OE}$ .

As a result, we may associate with any three collinear points  $A$ ,  $B$ , and  $C$  of an affine space, excepting only the case  $A = B = C$ , a scalar  $\alpha$ , called the *affine ratio* of the points  $A$ ,  $B$ , and  $C$  and denoted by  $(A, B, C)$ . This is accomplished as follows. If  $A \neq B$ , then  $\alpha$  is uniquely determined by the relationship  $\overrightarrow{AC} = \alpha \overrightarrow{AB}$ . In particular,  $\alpha = 1$  if  $B = C$ , and  $\alpha = 0$  if  $A = C$ . If  $A = B \neq C$ , then we take  $\alpha = \infty$ . And if all three points  $A$ ,  $B$ , and  $C$  coincide, then their affine ratio  $(A, B, C)$  is undefined.

Using the concept of oriented length of a line segment, we can write the affine ratio of three points using the following formula:

$$(A, B, C) = \frac{AC}{AB}, \quad (8.8)$$

where  $AB$  denotes the *signed* length of  $AB$ , that is,  $AB = |AB|$  if the point  $A$  lies to the left of  $B$ , and  $AB = -|AB|$  if the point  $A$  lies to the right of  $B$ . Here, of course, in formula (8.8), we assume that  $a/0 = \infty$  for every  $a \neq 0$ .

For the remainder of this section, we shall assume that  $V$  is a *real* affine space.

In this case, obviously, the numbers  $\alpha$  from relationship (8.7) corresponding to the points of the line  $P$  are real, and the relationship  $\alpha < \gamma < \beta$  between numbers on the real line carries over to the corresponding points of the line  $P \subset V$ . If these numbers  $\alpha$ ,  $\beta$ , and  $\gamma$  correspond to the points  $A$ ,  $B$ , and  $C$ , then we say that the point  $C$  *lies between* the points  $A$  and  $B$ .

Despite the fact that the relationship  $A \mapsto \alpha$  defined by formula (8.7) itself depends on the choice of distinct points  $O$  and  $E$  on the line, the property of point  $C$  that it lie between  $A$  and  $B$  does not depend on that choice (although with a different choice of  $O$  and  $E$ , the order of the points  $A$  and  $B$  might, of course, change). Indeed, it is easy to verify that by replacing the point  $O$  by  $O'$ , to each of the numbers  $\alpha$ ,  $\beta$ , and  $\gamma$  is added one and the same term  $\lambda$  corresponding to the vector  $\overrightarrow{OO'}$ , and

in replacing the point  $E$  by  $E'$ , each of the numbers  $\alpha$ ,  $\beta$ , and  $\gamma$  is multiplied by one and the same number  $\mu \neq 0$  such that  $\overrightarrow{OE} = \mu \overrightarrow{OE'}$ . For both operations, the relationship  $\alpha < \gamma < \beta$  for the point  $C$  and pair of points  $A$  and  $B$  is unchanged, except that the numbers  $\alpha$  and  $\beta$  in this inequality may exchange places (if they are multiplied by  $\mu < 0$ ).

The property of a point  $C$  to lie between  $A$  and  $B$  is related to the affine ratio for three collinear points introduced above. Namely, it is obvious that in the case of a real space, the inequality  $(C, A, B) < 0$  is satisfied if and only if the point  $C$  lies between  $A$  and  $B$ .

**Definition 8.17** The collection of all points on the line passing through the points  $A$  and  $B$  that lie between  $A$  and  $B$  together with  $A$  and  $B$  themselves is called the *segment* joining the points  $A$  and  $B$  and is denoted by  $[A, B]$ . Here the points  $A$  and  $B$  are called the *endpoints* of the segment, and by definition, they belong to it.

Thus the segment is determined by two points  $A$  and  $B$ , but not by their order, that is, by definition  $[B, A] = [A, B]$ .

**Definition 8.18** A set  $M \subset V$  is said to be *convex* if for every pair of points  $A, B \in M$ , the set  $M$  also contains the segment  $[A, B]$ .

The notion of convexity is related to the partition of an affine space  $V$  by a hyperplane  $V'$  into two half-spaces, in analogy with the partition of a vector space into two half-spaces constructed in Sect. 3.2. In order to define this partition, let us denote by  $L' \subset L$  the hyperplane corresponding to  $V'$ , and let us consider the partition  $L \setminus L' = L^+ \cup L^-$  introduced earlier, choose an arbitrary point  $O' \in V'$ , and for a point  $A \in V \setminus V'$ , state that  $A \in V^+$  or  $A \in V^-$  depending on the half-space ( $L^+$  or  $L^-$ ) to which the vector  $\overrightarrow{O'A}$  belongs.

A simple verification shows that the subsets  $V^+$  and  $V^-$  thus obtained depend only on the half-spaces  $L^+$  and  $L^-$  and not on the choice of point  $O' \in V'$ . Obviously,  $V \setminus V' = V^+ \cup V^-$  and  $V^+ \cap V^- = \emptyset$ .

**Theorem 8.19** *The sets  $V^+$  and  $V^-$  are convex, but the entire set  $V \setminus V'$  is not.*

*Proof* Let us begin by verifying the assertion about the set  $V^+$ . Let  $A, B \in V^+$ . This implies that the vectors  $\mathbf{x} = \overrightarrow{O'A}$  and  $\mathbf{y} = \overrightarrow{O'B}$  belong to the half-space  $L^+$ , that is, they can be expressed in the form

$$\mathbf{x} = \alpha \mathbf{e} + \mathbf{u}, \quad \mathbf{y} = \beta \mathbf{e} + \mathbf{v}, \quad \alpha, \beta > 0, \mathbf{u}, \mathbf{v} \in L', \quad (8.9)$$

for some fixed vector  $\mathbf{e} \notin L'$ . Let us consider the vector  $\mathbf{z} = \overrightarrow{O'C}$  and write it in the form

$$\mathbf{z} = \gamma \mathbf{e} + \mathbf{w}, \quad \mathbf{w} \in L'. \quad (8.10)$$

Assuming that the point  $C$  lies between  $A$  and  $B$ , let us prove that  $z \in L^+$ , that is, that  $\gamma > 0$ . The given condition, that the point  $C$  lies between  $A$  and  $B$ , can be written with the help of an association between the points on the line passing through  $A$  and  $B$  and the numbers that are the coordinates in the frame of reference  $(O; \overrightarrow{OE})$  according to formula (8.7). Although this association depends on the choice of points  $O$  and  $E$ , the property itself of “lying between,” as we have seen, does not depend on this choice. Therefore, we may choose  $O = A$  and  $E = B$ . Then in our frame of reference, the point  $A$  has coordinate 0, and the point  $B$  has coordinate 1. Let  $C$  have coordinate  $\lambda$ . Since  $C \in [A, B]$ , it follows that  $0 \leq \lambda \leq 1$ . By definition,  $\overrightarrow{AC} = \lambda \overrightarrow{AB}$ . But from the fact that

$$\overrightarrow{AC} = \overrightarrow{AO'} + \overrightarrow{O'C} = z - x, \quad \overrightarrow{AB} = \overrightarrow{AO'} + \overrightarrow{O'B} = y - x,$$

we obtain the equality  $z - x = \lambda(y - x)$ , or equivalently, the equality

$$z = (1 - \lambda)x + \lambda y.$$

Using formulas (8.9) and (8.10), we obtain from the last equality the relationship  $\gamma = (1 - \lambda)\alpha + \lambda\beta$ , from which, taking into account the inequalities  $\alpha > 0$ ,  $\beta > 0$ , and  $0 \leq \lambda \leq 1$ , it follows that  $\gamma > 0$ .

The convexity of the set  $V^-$  is proved in exactly the same way.

We shall prove, finally, that the set  $V \setminus V'$  is not convex. In view of the convexity of  $V^+$  and  $V^-$ , of interest to us is only the case in which the points  $A$  and  $B$  lie in different half-spaces, for example,  $A \in V^+$  and  $B \in V^-$  (or conversely,  $A \in V^-$  and  $B \in V^+$ , but this case is completely analogous). The condition  $A \in V^+$  and  $B \in V^-$  means that in formulas (8.9), we have  $\alpha > 0$  and  $\beta < 0$ . In analogy to what has gone before, for an arbitrary point  $C \in [A, B]$ , let us construct the vector  $z$  as was done in (8.10), and thus obtain the equality  $\gamma = (1 - \lambda)\alpha + \lambda\beta$ . If the numbers  $\alpha$  and  $\beta$  are of opposite sign, an elementary computation shows that there always exists a number  $\lambda \in [0, 1]$  such that  $(1 - \lambda)\alpha + \lambda\beta = 0$ , and this yields that  $C \in [A, B]$ . Thus the theorem is proved in its entirety.  $\square$

Thus the set  $V^+$  is characterized by the property that every pair of its points are connected by a segment lying entirely within it. This holds as well for the set  $V^-$ . At the same time, *no* two points  $A \in V^+$  and  $B \in V^-$  can be joined by a segment that does not intersect the hyperplane  $V'$ . This consideration gives another definition of the partition  $V \setminus V' = V^+ \cup V^-$ , one that does not appeal to vector spaces.

Let us consider the sequence of subspaces

$$V_0 \subset V_1 \subset V_2 \subset \cdots \subset V_n = V, \quad \dim V_i = i. \quad (8.11)$$

From the last condition, it follows that  $V_{i-1}$  is a hyperplane in  $V_i$ , and this implies that the partition defined by  $V_i \setminus V_{i-1} = V_i^+ \cup V_i^-$  is the partition introduced above.

A pair of half-spaces  $(V_{i-1}, V_i)$  is said to be *directed* if it is indicated which of two convex subsets of the set  $V_i \setminus V_{i-1}$  we denote by  $V_i^+$ , and which by  $V_i^-$ . The

sequence of subspaces (8.11) is called a *flag* if each pair  $(V_{i-1}, V_i)$  is directed. We note that in a flag defined by the sequence (8.11), the subspace  $V_0$  has dimension 0, that is, it consists of a single point. This point is called the *center* of the flag.

### 8.3 Affine Transformations

**Definition 8.20** An *affine transformation* of an affine space  $(V, L)$  into another affine space  $(V', L')$  is a pair of mappings

$$f : V \rightarrow V', \quad \mathcal{F} : L \rightarrow L',$$

satisfying the following two conditions:

- (1) The mapping  $\mathcal{F} : L \rightarrow L'$  is a linear transformation of vector spaces  $L \rightarrow L'$ .
- (2) For every pair of points  $A, B \in V$ , we have the equality

$$\overrightarrow{f(A)f(B)} = \mathcal{F}(\overrightarrow{AB}).$$

Condition (2) means that the linear transformation  $\mathcal{F}$  is determined by the mapping  $f$ . It is called the *linear part* of the mapping  $f$  and is denoted by  $\Lambda(f)$ . In the sequel we shall, as a rule, indicate only the mapping  $f : V \rightarrow V'$ , since the linear part  $\mathcal{F}$  is uniquely determined by it, and we shall view the affine transformation as a mapping from  $V$  to  $V'$ .

**Theorem 8.21** *Affine transformations possess the following properties:*

- (a) *The composition of two affine transformations  $f$  and  $g$  is again an affine transformation, which we denote by  $gf$ . Here  $\Lambda(gf) = \Lambda(g)\Lambda(f)$ .*
- (b) *An affine transformation  $f$  is bijective if and only if the linear transformation  $\Lambda(f)$  is bijective. In this case, the inverse transformation  $f^{-1}$  is also an affine transformation, and  $\Lambda(f^{-1}) = \Lambda(f)^{-1}$ .*
- (c) *If  $f = e$ , the identity transformation, then  $\Lambda(f) = \mathcal{E}$ .*

*Proof* All these assertions are proved by direct verification.

(a) Let  $(V, L)$ ,  $(V', L')$ , and  $(V'', L'')$  be affine spaces. Let us consider the affine transformation  $f : V \rightarrow V'$  with linear part  $\mathcal{F} = \Lambda(f)$  and another affine transformation  $g : V' \rightarrow V''$  with linear part  $\mathcal{G} = \Lambda(g)$ . We shall denote the composition of  $f$  and  $g$  by  $h$ , and the composition of  $\mathcal{F}$  and  $\mathcal{G}$  by  $\mathcal{H}$ . Then by the definition of the composition of arbitrary mappings of sets, we have  $h : V \rightarrow V''$  and  $\mathcal{H} : L \rightarrow L''$ , and moreover, we know that  $\mathcal{H}$  is a linear transformation. Thus we must show that every pair of points  $A, B \in V$  satisfies the equality  $\overrightarrow{h(A)h(B)} = \mathcal{H}(\overrightarrow{AB})$ . But since by definition, we have the equalities

$$\overrightarrow{f(A)f(B)} = \mathcal{F}(\overrightarrow{AB}), \quad \overrightarrow{g(A')g(B')} = \mathcal{G}(\overrightarrow{A'B'})$$

for arbitrary points  $A, B \in V$  and  $A', B' \in V'$ , it follows that

$$\overrightarrow{h(A)h(B)} = \overrightarrow{g(f(A))g(f(B))} = \overrightarrow{g(f(A)f(B))} = \overrightarrow{g(\mathcal{F}(\overrightarrow{AB}))} = \mathcal{H}(\overrightarrow{AB}).$$

The proofs of assertions (b) and (c) are just as straightforward.  $\square$

Let us give some examples of affine transformations.

**Example 8.22** For affine spaces  $(L, L)$  and  $(L', L')$ , a linear transformation  $f = \mathcal{F} : L \rightarrow L'$  is affine, and moreover, it is obvious that  $\Lambda(f) = \mathcal{F}$ .

In the sequel, we shall frequently encounter affine transformations in which the affine spaces  $V$  and  $V'$  coincide (and this also applies to the spaces of vectors  $L$  and  $L'$ ). We shall call such an affine transformation of a space  $V$  an affine transformation of the space *into itself*.

**Example 8.23** A translation  $\mathcal{T}_a$  by an arbitrary vector  $a \in L$  is an affine transformation of the space  $V$  into itself. It follows from the definition of translation that  $\Lambda(\mathcal{T}_a) = \mathcal{E}$ . Conversely, every affine transformation whose linear part is equal to  $\mathcal{E}$  is a translation. Indeed, by the definition of an affine transformation, the condition  $\Lambda(f) = \mathcal{E}$  implies that  $\overrightarrow{f(A)f(B)} = \overrightarrow{AB}$ . Recalling Remark 8.2 and Fig. 8.1, we see that from this assertion follows the equality  $\overrightarrow{Af(A)} = \overrightarrow{Bf(B)}$ , which implies that  $f = \mathcal{T}_a$ , where the vector  $a$  is equal to  $\overrightarrow{Af(A)}$  for some (any) point  $A$  of the space  $V$ .

The same reasoning allows us to obtain a more general result.

**Theorem 8.24** If affine transformations  $f : V \rightarrow V'$  and  $g : V \rightarrow V'$  have identical linear parts, then they differ only by a translation, that is, there exists a vector  $a \in L'$  such that  $g = \mathcal{T}_a f$ .

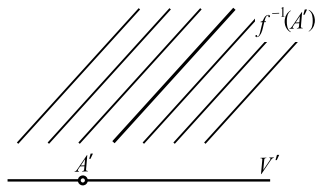
*Proof* By definition, the equality  $\Lambda(f) = \Lambda(g)$  implies that  $\overrightarrow{f(A)f(B)} = \overrightarrow{g(A)g(B)}$  for every pair of points  $A, B \in V$ . From this, the equality

$$\overrightarrow{f(A)g(A)} = \overrightarrow{f(B)g(B)} \quad (8.12)$$

clearly follows. As in Example 8.23, this reasoning is based on Remark 8.2. The relationship (8.12) implies that the vector  $\overrightarrow{f(A)g(A)}$  does not depend on the choice of the point  $A$ . We shall denote this vector by  $a$ . Then by the definition of translation,  $g(A) = \mathcal{T}_a(f(A))$  for every point  $A \in V$ , which completes the proof of the theorem.  $\square$

**Definition 8.25** Let  $V' \subset V$  be a subspace of the affine space  $V$ . An affine transformation  $f : V \rightarrow V'$  is said to be a *projection* onto the subspace  $V'$  if  $f(V) = V'$  and the restriction of  $f$  to  $V'$  is the identity transformation.

**Fig. 8.4** Fibers of a projection



**Theorem 8.26** If  $f : V \rightarrow V'$  is a projection onto the subspace  $V' \subset V$ , then the preimage  $f^{-1}(A')$  of an arbitrary point  $A' \in V'$  is an affine subspace of  $V$  of dimension  $\dim V - \dim V'$ . For distinct points  $A', A'' \in V'$ , the subspaces  $f^{-1}(A')$  and  $f^{-1}(A'')$  are parallel.

*Proof* Let  $\mathcal{F} = \Lambda(f)$ . Then  $\mathcal{F} : L \rightarrow L'$  is a linear transformation, where  $L$  and  $L'$  are the respective spaces of vectors of the affine spaces  $V$  and  $V'$ . Let us consider an arbitrary point  $A' \in V'$  and points  $P, Q \in f^{-1}(A')$ , that is,  $f(P) = f(Q) = A'$ . Then the vector  $\overrightarrow{f(P)f(Q)}$  is equal to  $\mathbf{0}$ , whence by the definition of an affine transformation, we obtain that  $\overrightarrow{f(P)f(Q)} = \mathcal{F}(\overrightarrow{PQ}) = \mathbf{0}$ , that is, the vector  $\overrightarrow{PQ}$  is in the kernel of the linear transformation  $\mathcal{F}$ , which, as we know, is a subspace of  $L$ .

Conversely, if  $P \in f^{-1}(A')$  and the vector  $x$  is in the kernel of the transformation  $\mathcal{F}$ , that is,  $\mathcal{F}(x) = \mathbf{0}$ , then there exists a point  $Q \in V$  for which  $x = \overrightarrow{PQ}$ . Then  $f(P) = f(Q)$  and  $Q \in f^{-1}(A')$ . By definition, an arbitrary vector  $x = \overrightarrow{A'B'} \in L'$  can be represented in the form  $\mathcal{F}(\overrightarrow{PQ})$ , where  $f(P) = A'$  and  $f(Q) = B'$ . This means that the image of the transformation  $\mathcal{F}$  coincides with the entire space  $L'$ , whence by Theorem 3.72, we obtain

$$\dim f^{-1}(A') = \dim \mathcal{F}^{-1}(\mathbf{0}) = \dim L - \dim L' = \dim V - \dim V',$$

since  $\mathcal{F}^{-1}(\mathbf{0})$  is the kernel of the transformation  $\mathcal{F}$ , and the number  $\dim L'$  is equal to its rank; see Fig. 8.4. We have already proved that for every point  $A' \in V'$ , the space of vectors of the affine space  $f^{-1}(A')$  coincides with  $\mathcal{F}^{-1}(\mathbf{0})$ . This completes the proof of the theorem.  $\square$

The subspaces  $f^{-1}(A')$  for the points  $A' \in V'$  are called *fibers* of the projection  $f : V \rightarrow V'$ ; see Fig. 8.4. If  $S' \subset V'$  is some subset (not necessarily a subspace), then its preimage, the set  $S = f^{-1}(S')$ , is called a *cylinder* in  $V$ .

**Definition 8.27** An affine transformation  $f : V \rightarrow V'$  is called an *isomorphism* if it is a bijection. Affine spaces  $V$  and  $V'$  in this case are said to be *isomorphic*.

By assertion (b) of Theorem 8.21, the condition of a transformation  $f : V \rightarrow V'$  being a bijection is equivalent to the bijectivity of the linear transformation  $\Lambda(f) : L \rightarrow L'$  of the corresponding spaces of vectors  $L$  and  $L'$ . Thus affine spaces  $V$  and  $V'$  are isomorphic if and only if the corresponding spaces of vectors  $L$  and  $L'$  are isomorphic. As shown in Sect. 3.5, vector spaces  $L$  and  $L'$  are isomorphic if and

only if  $\dim L = \dim L'$ , and in this situation every nonsingular linear transformation  $L \rightarrow L'$  is an isomorphism. This yields the following assertion: affine spaces  $V$  and  $V'$  are isomorphic if and only if  $\dim V = \dim V'$ . Here every affine transformation  $f : V \rightarrow V'$  whose linear part  $\Lambda(f)$  is nonsingular is an isomorphism between  $V$  and  $V'$ . We shall frequently call an affine transformation  $f$  with nonsingular linear part  $\Lambda(f)$  *nonsingular*.

From the definitions, we immediately obtain the following theorem.

**Theorem 8.28** *The affine ratio  $(A, B, C)$  of three collinear points does not change under a nonsingular affine transformation.*

*Proof* By definition, the affine ratio  $\alpha = (A, B, C)$  of three points  $A, B, C$  under the condition  $A \neq B$  is defined by the relationship

$$\overrightarrow{AC} = \alpha \overrightarrow{AB}. \quad (8.13)$$

Let  $f : V \rightarrow V$  be a nonsingular affine transformation and  $\mathcal{F} : L \rightarrow L$  its corresponding linear transformation. Then in view of the nondegeneracy of the transformation  $f$ , we have  $f(A) \neq f(B)$  and

$$\overrightarrow{f(A)f(C)} = \mathcal{F}(\overrightarrow{AC}), \quad \overrightarrow{f(A)f(B)} = \mathcal{F}(\overrightarrow{AB}),$$

and  $\beta = (f(A), f(B), f(C))$  is defined by the equality  $\overrightarrow{f(A)f(C)} = \beta \overrightarrow{f(A)f(B)}$ , that is,

$$\mathcal{F}(\overrightarrow{AC}) = \beta \mathcal{F}(\overrightarrow{AB}). \quad (8.14)$$

Applying the transformation  $\mathcal{F}$  to both sides of equality (8.13), we obtain  $\mathcal{F}(\overrightarrow{AC}) = \alpha \mathcal{F}(\overrightarrow{AB})$ , whence taking into account equality (8.14), it follows that  $\beta = \alpha$ . In the case that  $A = B \neq C$ , we obtain, in view of the nonsingularity of  $f$ , the analogous relationship  $f(A) = f(B) \neq f(C)$ , from which we have  $(A, B, C) = \infty$  and  $(f(A), f(B), f(C)) = \infty$ .  $\square$

**Example 8.29** Every affine space  $(V, L)$  is isomorphic to the space  $(L, L)$ . Indeed, let us choose in the set  $V$  an arbitrary point  $O$  and define the mapping  $f : V \rightarrow L$  in such a way that  $f(A) = \overrightarrow{OA}$ . It is obvious, by the definition of affine space, that the mapping  $f$  is an isomorphism.

Let us note that the situation here is similar to that of an isomorphism of a vector space  $L$  and the dual space  $L^*$ . In one case, the isomorphism requires the choice of a basis of  $L$ , while in the other, it is the choice of a point  $O$  in  $V$ .

Let  $f : V \rightarrow V'$  be an affine transformation of affine spaces  $(V, L)$  and  $(V', L')$ . Let us consider isomorphisms  $\varphi : V \rightarrow L$  and  $\varphi' : V' \rightarrow L'$ , defined, as in Example 8.29, by the selection of certain points  $O \in V$  and  $O' \in V'$ . We have the map-



pings

$$\begin{array}{ccc}
 V & \xrightarrow{f} & V' \\
 \varphi \downarrow & & \downarrow \varphi' \\
 L & \xrightarrow{\mathcal{F}} & L'
 \end{array} \tag{8.15}$$

where  $\mathcal{F} = \Lambda(f)$ . Here, generally speaking, we cannot assert that  $\mathcal{F}\varphi = \varphi'f$ , but nevertheless, these mappings are closely related. For an arbitrary point  $A \in V$ , we have by construction that  $\varphi(A) = \overrightarrow{OA}$  and  $\mathcal{F}(\varphi(A)) = \mathcal{F}(\overrightarrow{OA}) = \overrightarrow{f(O)f(A)}$ . In just the same way,  $\varphi'(f(A)) = \overrightarrow{O'f(A)}$ . Finally,  $\overrightarrow{O'f(A)} = \overrightarrow{O'f(O)} + \overrightarrow{f(O)f(A)}$ . Combining these relationships, we obtain

$$\varphi'f = \mathcal{T}_b\mathcal{F}, \quad \text{where } b = \overrightarrow{O'f(O)}. \tag{8.16}$$

Relationship (8.16) allows us to write down the action of affine transformations in coordinate form. To do so, we choose frames of reference  $(O; e_1, \dots, e_n)$  and  $(O'; e'_1, \dots, e'_m)$ , where  $n = \dim V$  and  $m = \dim V'$ , in the spaces  $V$  and  $V'$ . Then the coordinates of the point  $A$  in the chosen frame of reference are the coordinates of the vector  $\overrightarrow{OA} = \varphi(A)$  in the basis  $e_1, \dots, e_n$ . Likewise, the coordinates of the point  $f(A)$  are the coordinates of the vector  $\overrightarrow{O'f(A)} = \varphi'(f(A))$  in the basis  $e'_1, \dots, e'_m$ . Let us make use of relationship (8.16). Suppose the coordinates of the vector  $\overrightarrow{OA}$  in the basis  $e_1, \dots, e_n$  are  $(\alpha_1, \dots, \alpha_n)$ , the coordinates of the vector  $\overrightarrow{O'f(A)}$  in the basis  $e'_1, \dots, e'_m$  are  $(\alpha'_1, \dots, \alpha'_m)$ , and the matrix of the linear transformation  $\mathcal{F}$  in these bases is  $F = (f_{ij})$ . Setting the coordinates of the vector  $b$  from formula (8.16) in the basis  $e'_1, \dots, e'_m$  equal to  $(\beta_1, \dots, \beta_m)$ , we obtain

$$\alpha'_i = \sum_{j=1}^n f_{ij}\alpha_j + \beta_i, \quad i = 1, \dots, m. \tag{8.17}$$

Using the standard notation for column vectors

$$[\alpha] = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}, \quad [\alpha'] = \begin{pmatrix} \alpha'_1 \\ \vdots \\ \alpha'_m \end{pmatrix}, \quad [\beta] = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix},$$

we may rewrite formula (8.17) in the form

$$[\alpha'] = F[\alpha] + [\beta]. \tag{8.18}$$

The most frequent case that we shall encounter in the sequel is that of transformations of an affine space  $V$  into itself. Let us assume that the mapping  $f : V \rightarrow V$  has a *fixed point*  $O$ , that is, for the point  $O \in V$ , we have  $f(O) = O$ . Then the transformation  $f$  can be identified with its linear part, that is, if by the choice of affine

space  $V$ , the frame of reference  $(O; \mathbf{e}_1, \dots, \mathbf{e}_n)$  with fixed point  $O$  identifies  $V$  with the vector space  $L$ , then the mapping  $\overrightarrow{f}$  is identified with its linear part  $\mathcal{F} = \Lambda(f)$ . Here  $f(O) = O$  and  $\overrightarrow{Of(A)} = \mathcal{F}(\overrightarrow{OA})$  for every point  $A \in V$ .

We shall call such affine transformations of a space  $V$  into itself *linear* (we note that this notion depends on the choice of point  $O \in V$  that  $f$  maps to itself). If for an arbitrary affine transformation  $f$  we define  $f_0 = \mathcal{T}_a^{-1} f$ , where the vector  $\mathbf{a}$  is equal to  $\overrightarrow{Of(O)}$ , then  $f_0$  will be a linear transformation, and we obtain the representation

$$f = \mathcal{T}_a f_0. \quad (8.19)$$

It is obvious that a nonsingular affine transformation of the space  $(V, L)$  takes each frame of reference  $(O; \mathbf{e}_1, \dots, \mathbf{e}_n)$  into some other frame of reference. This implies that if  $f(O) = O'$  and  $\Lambda(f)(\mathbf{e}_i) = \mathbf{e}'_i$ , then  $(O'; \mathbf{e}'_1, \dots, \mathbf{e}'_n)$  is also a frame of reference. Conversely, if the transformation  $f$  takes some frame of reference to another frame of reference, then it is nonsingular.

From the representation (8.19) we obtain the following result.

If we are given a frame of reference  $(O; \mathbf{e}_1, \dots, \mathbf{e}_n)$ , an arbitrary point  $O'$ , and vectors  $\mathbf{a}_1, \dots, \mathbf{a}_n$  in  $L$ , then there exists (and it is unique) an affine transformation  $f$  mapping  $O$  to  $O'$  such that  $\Lambda(f)(\mathbf{e}_i) = \mathbf{a}_i$  for all  $i = 1, \dots, n$ . To prove this, we set  $\mathbf{a}$  equal to  $\overrightarrow{OO'}$  in representation (8.19), and for  $f_0$ , we take a linear transformation of the vector space  $L$  into itself such that  $f_0(\mathbf{e}_i) = \mathbf{a}_i$  for all  $i = 1, \dots, n$ . It is obvious that the affine transformation  $f$  thus constructed satisfies the requisite conditions. Its uniqueness follows from the representation (8.19) and from the fact that the vectors  $\mathbf{e}_1, \dots, \mathbf{e}_n$  form a basis of  $L$ .

The following reformulation of this statement is obvious: if we are given  $n + 1$  independent points  $A_0, A_1, \dots, A_n$  of an  $n$ -dimensional affine space  $V$  and an additional arbitrary  $n + 1$  points  $B_0, B_1, \dots, B_n$ , then there exists (and it is unique) an affine transformation  $f: V \rightarrow V$  such that  $f(A_i) = B_i$  for all  $i = 0, 1, \dots, n$ .

In the sequel, it will be useful to know about the dependence of the vector  $\mathbf{a}$  in representation (8.19) on the choice of point  $O$  (on its choice also depends the transformation  $f_0$  of the space  $V$ , but as a transformation of a vector space  $L$ , it coincides with  $\Lambda(f)$ ). Let us set  $\overrightarrow{OO'} = \mathbf{c}$ . Then for a new choice of  $O'$  as fixed point, we have, similar to (8.19), the representation

$$f = \mathcal{T}_{a'} f'_0, \quad (8.20)$$

where  $f'_0(O') = O'$  and the vector  $\mathbf{a}'$  is equal to  $\overrightarrow{O'f(O')}$ . By well-known rules, we have

$$\begin{aligned} \mathbf{a}' &= \overrightarrow{O'f(O')} = \overrightarrow{O'O} + \overrightarrow{Of(O')}, \\ \overrightarrow{Of(O')} &= \overrightarrow{Of(O)} + \overrightarrow{f(O)f(O')} = \mathbf{a} + \mathcal{F}(\mathbf{c}). \end{aligned}$$

Since  $\overrightarrow{O'O} = -\overrightarrow{OO'}$ , we obtain that the vectors  $\mathbf{a}$  and  $\mathbf{a}'$  in representations (8.19) and (8.20) are related by

$$\mathbf{a}' = \mathbf{a} + \mathcal{F}(\mathbf{c}) - \mathbf{c}, \quad \text{where } \mathbf{c} = \overrightarrow{OO'}. \quad (8.21)$$

Let us choose a frame of reference in the affine space  $(V, L)$ . Let us recall that it is written in the form  $(O; \mathbf{e}_1, \dots, \mathbf{e}_n)$  or  $(O; A_1, \dots, A_n)$ , where  $\mathbf{e}_i = \overrightarrow{OA_i}$ . Let  $f$  be a nonsingular transformation of  $V$  into itself, and let it map the frame of reference  $(O; \mathbf{e}_1, \dots, \mathbf{e}_n)$  to  $(O'; \mathbf{e}'_1, \dots, \mathbf{e}'_n)$ . If  $\mathbf{e}'_i = \overrightarrow{O'A'_i}$ , then this implies that  $f(O) = O'$  and  $f(A_i) = A'_i$  for  $i = 1, \dots, n$ .

Let the point  $A \in V$  have coordinates  $(\alpha_1, \dots, \alpha_n)$  relative to the frame of reference  $(O; A_1, \dots, A_n)$ . This means that the vector  $\overrightarrow{OA}$  is equal to  $\alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n$ . Then the point  $f(A)$  determines the vector  $\overrightarrow{f(O)f(A)}$ , that is,  $\mathcal{F}(\overrightarrow{OA})$ . And this vector obviously has, in the basis  $\mathbf{e}'_1, \dots, \mathbf{e}'_n$ , the same coordinates as the vector  $\overrightarrow{OA}$  in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , since by definition,  $\mathbf{e}'_i = \mathcal{F}(\mathbf{e}_i)$ . Thus the affine transformation  $f$  is defined by the fact that the point  $A$  is mapped to a different point  $f(A)$  having in the frame of reference  $(O', \mathbf{e}'_1, \dots, \mathbf{e}'_n)$  the same coordinates as the point  $A$  had in the frame of reference  $(O; \mathbf{e}_1, \dots, \mathbf{e}_n)$ .

**Definition 8.30** Two subsets  $S$  and  $S'$  of an affine space  $V$  are said to be *affinely equivalent* if there exists a nonsingular affine transformation  $f: V \rightarrow V$  such that  $f(S) = S'$ .

The previous reasoning shows that this definition is equivalent to saying that in the space  $V$ , there exist two frames of reference  $(O; \mathbf{e}_1, \dots, \mathbf{e}_n)$  and  $(O'; \mathbf{e}'_1, \dots, \mathbf{e}'_n)$  such that all points of the set  $S$  have the same coordinates with respect to the first frame of reference as the points of the set  $S'$  have with respect to the second.

In the case of real affine spaces, the definition of affine transformations by formulas (8.17) and (8.18) makes it possible to apply to them Theorem 4.39 on proper and improper linear transformations.

**Definition 8.31** A nonsingular affine transformation of a real affine space  $V$  to itself is said to be *proper* if its linear part is a proper transformation of the vector space. Otherwise, it is called *improper*.

Thus by this definition, we consider translations to be proper transformations. A bit later, we shall provide a more meaningful justification for this definition.

By the given definition of affine transformation, whether  $f$  is proper or improper depends on the sign of the determinant of the matrix  $F = (f_{ij})$  in formulas (8.17), (8.18). We observe that this concept relates only to nonsingular transformations  $V$ , since in formulas (8.17) and (8.18), we must have  $m = n$ .

In order to formulate an analogue to Theorem 4.39, we should refine the sense of the assertion about the fact that the family  $g(t)$  of affine transformations depends

continuously on the parameter  $t$ . By this, we shall understand that for  $g(t)$ , in the formula

$$\alpha'_i = \sum_{j=1}^n g_{ij}(t)\alpha_j + \beta_i(t), \quad i = 1, \dots, n, \quad (8.22)$$

analogous to (8.17), written in some (arbitrarily chosen) frame of reference of the space  $V$ , all coefficients  $g_{ij}(t)$  and  $\beta_i(t)$  depend continuously on  $t$ . In particular, if  $G(t) = (g_{ij}(t))$  is a matrix of the linear part of the affine transformation  $g(t)$ , then its determinant  $|G(t)|$  is a continuous function. From the properties of continuous functions, it follows that the determinant  $|G(t)|$  has the same sign at all points of the interval  $[0, 1]$ .

Thus we shall say that an affine transformation  $f$  is *continuously deformable* into  $h$  if there exists a family  $g(t)$  of continuous affine transformations, depending continuously on the parameter  $t \in [0, 1]$ , such that  $g(0) = f$  and  $g(1) = h$ . It is obvious that the property thus defined of affine transformations being continuously deformable into each other defines on the set of such transformations an equivalence relation, that is, it satisfies the properties of reflexivity, symmetry, and transitivity.

**Theorem 8.32** *Two nondegenerate affine transformations of a real affine space are continuously deformable into each other if and only if they are either both proper or both improper. In particular, a nonsingular affine transformation  $f$  is proper if and only if it is deformable into the identity.*

*Proof* Let us begin with the latter, more specific, assertion of the theorem. Let a nonsingular affine transformation  $f$  be continuously deformable into  $e$ . Then by symmetry, there exists a continuous family of nonsingular affine transformations  $g(t)$  with linear part  $\Lambda(g(t))$  such that  $g(0) = e$  and  $g(1) = f$ . For the transformation  $g(t)$ , let us write (8.22) in some frame of reference  $(O; e_1, \dots, e_n)$  of the space  $V$ . It is obvious that for the matrix  $G(t) = (g_{ij}(t))$ , we have the relationships  $G(0) = E$  and  $G(1) = F$ , where  $F$  is the matrix of the linear transformation  $\mathcal{F} = \Lambda(f)$  in the basis  $e_1, \dots, e_n$  of the space  $L$  and  $\beta_i(0) = 0$  for all  $i = 1, \dots, n$ . By the definition of continuous deformation, the determinant  $|G(t)|$  is nonzero for all  $t \in [0, 1]$ . Since  $|G(0)| = |E| = 1$ , it follows that  $|G(t)| > 0$  for all  $t \in [0, 1]$ , and in particular, for  $t = 1$ . And this means that  $|\Lambda(f)| = |G(1)| > 0$ . Thus the linear transformation  $\Lambda(f)$  is proper, and by definition, the affine transformation  $f$  is also proper.

Conversely, let  $f$  be a proper affine transformation. This means that the linear transformation  $\Lambda(f)$  is proper. Then by Theorem 4.39, the transformation  $\Lambda(f)$  is continuously deformable into the identity. Let  $\mathcal{G}(t)$  be a family of linear transformations such that  $\mathcal{G}(0) = \mathcal{E}$  and  $\mathcal{G}(1) = \Lambda(f)$ , given in some basis  $e_1, \dots, e_n$  of the space  $L$  by the formula

$$\alpha'_i = \sum_{j=1}^n g_{ij}(t)\alpha_j, \quad i = 1, \dots, n, \quad (8.23)$$

where  $g_{ij}(t)$  are continuous functions, the matrix  $G(t) = (g_{ij}(t))$  is nonsingular for all  $t \in [0, 1]$ , and we have the equalities  $G(0) = E$ ,  $G(t) = F$ , where  $F$  is the matrix of the transformation  $\Lambda(f)$  in the same basis  $e_1, \dots, e_n$ .

Let us consider the family  $g(t)$  of affine transformations given in the frame of reference  $(O; e_1, \dots, e_n)$  by the formula

$$\alpha'_i = \sum_{j=1}^n g_{ij}(t)\alpha_j + \beta_i t, \quad i = 1, \dots, n,$$

in which the coefficients of  $g_{ij}(t)$  are taken from formula (8.23), while the coefficients  $\beta_i$  are from formula (8.17) for the transformation  $f$  in the same frame of reference  $(O; e_1, \dots, e_n)$ . Since  $g(0) = \mathcal{E}$  and  $g(1) = \Lambda(f)$ , it is obvious that  $g(0) = e$  and  $g(1) = f$ , and moreover,  $|G(t)| > 0$  for all  $t \in [0, 1]$ , that is, the transformation  $g(t)$  is nonsingular for all  $t \in [0, 1]$ .

From this it follows by transitivity that every pair of proper affine transformations are continuously deformable into each other.

The case of improper affine transformations is handled completely analogously. It is necessary only to note that in all the arguments above, one must replace the identity transformation  $\mathcal{E}$  by some fixed improper linear transformation of the space  $L$ .  $\square$

Theorem 8.32 shows that analogously to real vector spaces, in every real affine space there exist two orientations, from which we may select arbitrarily whichever one we wish.

## 8.4 Affine Euclidean Spaces and Motions

**Definition 8.33** An affine space  $(V, L)$  is called an *affine Euclidean space* if the vector space  $L$  is a Euclidean space.

This means that for every pair of vectors  $x, y \in L$  there is defined a scalar product  $(x, y)$  satisfying the conditions enumerated in Sect. 7.1. In particular,  $(x, x) \geq 0$  for all  $x \in L$  and there is a definition of the length  $|x| = \sqrt{(x, x)}$  of a vector  $x$ . Since every pair of points  $A, B \in V$  defines a vector  $\overrightarrow{AB} \in L$ , it follows that one can associate with every pair of points  $A$  and  $B$ , the number

$$r(A, B) = |\overrightarrow{AB}|,$$

called the *distance* between the points  $A$  and  $B$  in  $V$ . This notion of distance that we have introduced satisfies the conditions for a metric introduced on p. xvii:

- (1)  $r(A, B) > 0$  for  $A \neq B$  and  $r(A, A) = 0$ ;
- (2)  $r(A, B) = r(B, A)$  for every pair of points  $A$  and  $B$ ;

(3) for every three points  $A$ ,  $B$ , and  $C$ , the triangle inequality is satisfied:

$$r(A, C) \leq r(A, B) + r(B, C). \quad (8.24)$$

Properties (1) and (2) clearly follow from the properties of the scalar product. Let us prove inequality (8.24), a special case of which (for right triangles) was proved on p. 216. By definition, if  $\overrightarrow{AB} = \mathbf{x}$  and  $\overrightarrow{BC} = \mathbf{y}$ , then (8.24) is equivalent to the inequality

$$|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|. \quad (8.25)$$

Since there are nonnegative numbers on the left- and right-hand sides of (8.25), we can square both sides and obtain an equivalent inequality, which we shall prove:

$$|\mathbf{x} + \mathbf{y}|^2 \leq (|\mathbf{x}| + |\mathbf{y}|)^2. \quad (8.26)$$

Since

$$|\mathbf{x} + \mathbf{y}|^2 = (\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) = |\mathbf{x}|^2 + 2(\mathbf{x}, \mathbf{y}) + |\mathbf{y}|^2,$$

then after multiplying out the right-hand side of (8.26), we can rewrite this inequality in the form

$$|\mathbf{x}|^2 + 2(\mathbf{x}, \mathbf{y}) + |\mathbf{y}|^2 \leq |\mathbf{x}|^2 + 2|\mathbf{x}| \cdot |\mathbf{y}| + |\mathbf{y}|^2.$$

Subtracting like terms from the left- and right-hand sides, we arrive at the inequality

$$(\mathbf{x}, \mathbf{y}) \leq |\mathbf{x}| \cdot |\mathbf{y}|,$$

which is the Cauchy–Schwarz inequality (7.6).

Thus an affine Euclidean space is a metric space.

In Sect. 8.1, we defined a frame of reference of an affine space as a point  $O$  in  $V$  and a basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  in  $L$ . If our affine space  $(V, L)$  is a Euclidean space, and the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  is orthonormal, then the frame of reference  $(O; \mathbf{e}_1, \dots, \mathbf{e}_n)$  is also said to be *orthonormal*. We see that an orthonormal frame of reference can be associated with each point  $O \in V$ .

**Definition 8.34** A mapping  $g : V \rightarrow V$  of an affine Euclidean space  $V$  into itself is said to be a *motion* if it is an isometry of  $V$  as a metric space, that is, if it preserves distances between points. This means that for every pair of points  $A, B \in V$ , the following equality holds:

$$r(g(A), g(B)) = r(A, B). \quad (8.27)$$

Let us emphasize that in this definition, we are speaking about an arbitrary mapping  $g : V \rightarrow V$ , which in general, does not have to be an affine transformation. By the discussion presented on p. xxi, a mapping  $g : V \rightarrow V$  is a motion if its image  $g(V) = V$  also satisfies the condition (8.27) of preserving distances.

*Example 8.35* Let  $\mathbf{a}$  be a vector in the vector space  $L$  corresponding to the affine space  $V$ . Then the translation  $\mathcal{T}_{\mathbf{a}}$  is a motion. Indeed, by the definition of a translation, for every point  $A \in V$  we have the equality  $\mathcal{T}_{\mathbf{a}}(A) = B$ , where  $\overrightarrow{AB} = \mathbf{a}$ . If for some other point  $C$ , we have an analogous equality  $\mathcal{T}_{\mathbf{a}}(C) = D$ , then  $\overrightarrow{CD} = \mathbf{a}$ . By condition (2) in the definition of an affine space, we have the equality  $\overrightarrow{AB} = \overrightarrow{CD}$ , from which, by Remark 8.2, it follows that  $\overrightarrow{AC} = \overrightarrow{BD}$ . This means that  $|\overrightarrow{AC}| = |\overrightarrow{BD}|$ , or equivalently,  $r(A, C) = r(\mathcal{T}_{\mathbf{a}}(A), \mathcal{T}_{\mathbf{a}}(C))$ , as asserted.

*Example 8.36* Let us assume that the mapping  $g : V \rightarrow V$  has the fixed point  $O$ , that is, the point  $O \in V$  satisfies the equality  $g(O) = O$ . As we saw in Sect. 8.3, the choice of point  $O$  determines a bijective mapping  $V \rightarrow L$ , where  $L$  is the space of vectors of the affine space  $V$ . Here to a point  $A \in V$  corresponds the vector  $\overrightarrow{OA} \in L$ .

Thus the mapping  $g : V \rightarrow V$  defines a mapping  $\mathcal{G} : L \rightarrow L$  such that  $\mathcal{G}(\mathbf{0}) = \mathbf{0}$ . Let us emphasize that since we did not assume that the mapping  $g$  was an affine transformation, the mapping  $\mathcal{G}$ , in general, is not a linear transformation of the space  $L$ . Now let us check that if  $\mathcal{G}$  is a linear orthogonal transformation of the Euclidean space  $L$ , then  $g$  is a motion.

By definition, the transformation  $\mathcal{G}$  is defined by the condition  $\mathcal{G}(\overrightarrow{OA}) = \overrightarrow{Og(A)}$ . We must prove that  $g$  is a motion, that is, that for all pairs of points  $A$  and  $B$ , we have

$$|\overrightarrow{g(A)g(B)}| = |\overrightarrow{AB}|. \quad (8.28)$$

We have the equality  $\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA}$ , and we obtain that

$$\overrightarrow{g(A)g(B)} = \overrightarrow{g(A)O} + \overrightarrow{Og(B)} = \overrightarrow{Og(B)} - \overrightarrow{Og(A)},$$

and this vector, by the definition of the transformation  $\mathcal{G}$ , is equal to  $\mathcal{G}(\overrightarrow{OB}) - \mathcal{G}(\overrightarrow{OA})$ . In view of the fact that the transformation  $\mathcal{G}$  is assumed to be linear, this vector is equal to  $\mathcal{G}(\overrightarrow{OB} - \overrightarrow{OA})$ . But as we have seen,  $\overrightarrow{OB} - \overrightarrow{OA} = \overrightarrow{AB}$ , and this means that

$$\overrightarrow{g(A)g(B)} = \mathcal{G}(\overrightarrow{AB}).$$

From the orthogonality of the transformation  $\mathcal{G}$  it follows that  $|\mathcal{G}(\overrightarrow{AB})| = |\overrightarrow{AB}|$ . In combination with the previous relationships, this yields the required equality (8.28).

The concept of motion is the most natural mathematical abstraction corresponding to the idea of the displacement of a solid body in space. We may apply to the analysis of this all of the results obtained in the preceding chapters, on the basis of the following fundamental assertion.

**Theorem 8.37** *Every motion is an affine transformation.*

*Proof* Let  $f$  be a motion of the affine Euclidean space  $V$ . As a first step, let us choose in  $V$  an arbitrary point  $O$  and consider the vector  $\mathbf{a} = \overrightarrow{Of(O)}$  and mapping

$g = \mathcal{T}_{-a}f$  of the space  $V$  into itself. Here the product  $\mathcal{T}_{-a}f$ , as usual, denotes sequential application (composition) of the mappings  $f$  and  $\mathcal{T}_{-a}$ . Then  $O$  is a fixed point of the transformation  $g$ , that is,  $g(O) = O$ . Indeed,  $g(O) = \mathcal{T}_{-a}(f(O))$ , and by the definition of translation, the equality  $g(O) = O$  is equivalent to  $\overrightarrow{f(O)O} = -\mathbf{a}$ , and this clearly follows from the fact that  $\mathbf{a} = \overrightarrow{Of(O)}$ .

We now observe that the product (that is, the sequential application, or composition) of two motions  $g_1$  and  $g_2$  is also a motion; the verification of this follows at once from the definition. Since we know that  $\mathcal{T}_a$  is a motion (see Example 8.35), it follows that  $g$  is also a motion. We therefore obtain a representation of  $f$  in the form  $f = \mathcal{T}_a g$ , where  $g$  is a motion and  $g(O) = O$ . Thus as we saw in Example 8.36,  $g$  defines a mapping  $\mathcal{G}$  of the space  $L$  into itself. The main part of the proof consists in verifying that  $\mathcal{G}$  is a linear transformation.

We shall base this verification on the following simple proposition.

**Lemma 8.38** *Assume that we are given a mapping  $\mathcal{G}$  of a vector space  $L$  into itself and a basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of  $L$ . Let us set  $\mathcal{G}(\mathbf{e}_i) = \mathbf{e}'_i$ ,  $i = 1, \dots, n$ , and assume that for every vector*

$$\mathbf{x} = \alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n, \quad (8.29)$$

*its image*

$$\mathcal{G}(\mathbf{x}) = \alpha_1 \mathbf{e}'_1 + \dots + \alpha_n \mathbf{e}'_n \quad (8.30)$$

*has the same  $\alpha_1, \dots, \alpha_n$ . Then  $\mathcal{G}$  is a linear transformation.*

*Proof* We must verify two conditions that enter into the definition of a linear transformation:

- (a)  $\mathcal{G}(\mathbf{x} + \mathbf{y}) = \mathcal{G}(\mathbf{x}) + \mathcal{G}(\mathbf{y})$ ,
- (b)  $\mathcal{G}(\alpha \mathbf{x}) = \alpha \mathcal{G}(\mathbf{x})$ ,

for all vectors  $\mathbf{x}$  and  $\mathbf{y}$  and scalar  $\alpha$ .

The verification of this is trivial. (a) Let the vectors  $\mathbf{x}$  and  $\mathbf{y}$  be given by  $\mathbf{x} = \alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n$  and  $\mathbf{y} = \beta_1 \mathbf{e}_1 + \dots + \beta_n \mathbf{e}_n$ . Then their sum is given by

$$\mathbf{x} + \mathbf{y} = (\alpha_1 + \beta_1) \mathbf{e}_1 + \dots + (\alpha_n + \beta_n) \mathbf{e}_n.$$

On the other hand, by the condition of the lemma, we have

$$\begin{aligned} \mathcal{G}(\mathbf{x} + \mathbf{y}) &= (\alpha_1 + \beta_1) \mathbf{e}'_1 + \dots + (\alpha_n + \beta_n) \mathbf{e}'_n \\ &= (\alpha_1 \mathbf{e}'_1 + \dots + \alpha_n \mathbf{e}'_n) + (\beta_1 \mathbf{e}'_1 + \dots + \beta_n \mathbf{e}'_n) = \mathcal{G}(\mathbf{x}) + \mathcal{G}(\mathbf{y}). \end{aligned}$$

- (b) For the vector  $\mathbf{x} = \alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n$  and an arbitrary scalar  $\alpha$ , we have

$$\alpha \mathbf{x} = (\alpha \alpha_1) \mathbf{e}_1 + \dots + (\alpha \alpha_n) \mathbf{e}_n.$$

By the condition of the lemma,

$$\mathcal{G}(\alpha \mathbf{x}) = (\alpha \alpha_1) \mathbf{e}'_1 + \dots + (\alpha \alpha_n) \mathbf{e}'_n = \alpha (\alpha_1 \mathbf{e}'_1 + \dots + \alpha_n \mathbf{e}'_n) = \alpha \mathcal{G}(\mathbf{x}). \quad \square$$



We now return to the proof of Theorem 8.37. Let us verify that the above construction of the mapping  $\mathcal{G} : \mathbf{L} \rightarrow \mathbf{L}$  satisfies the condition of the lemma. To this end, let us first ascertain that it preserves the inner product in  $\mathbf{L}$ , that is, that for all vectors  $\mathbf{x}, \mathbf{y} \in \mathbf{L}$ , we have the equality

$$(\mathcal{G}(\mathbf{x}), \mathcal{G}(\mathbf{y})) = (\mathbf{x}, \mathbf{y}). \quad (8.31)$$

Let us recall that the property for the transformation  $g$  to be a motion can be formulated as the following condition on a transformation  $\mathcal{G}$  of a vector space  $\mathbf{L}$ :

$$|\mathcal{G}(\mathbf{x}) - \mathcal{G}(\mathbf{y})| = |\mathbf{x} - \mathbf{y}| \quad (8.32)$$

for all pairs of vectors  $\mathbf{x}$  and  $\mathbf{y}$ . Squaring both sides of equality (8.32), we obtain

$$|\mathcal{G}(\mathbf{x}) - \mathcal{G}(\mathbf{y})|^2 = |\mathbf{x} - \mathbf{y}|^2. \quad (8.33)$$

Since  $\mathbf{x}$  and  $\mathbf{y}$  are vectors in the Euclidean space  $\mathbf{L}$ , we have

$$\begin{aligned} |\mathbf{x} - \mathbf{y}|^2 &= |\mathbf{x}|^2 - 2(\mathbf{x}, \mathbf{y}) + |\mathbf{y}|^2, \\ |\mathcal{G}(\mathbf{x}) - \mathcal{G}(\mathbf{y})|^2 &= |\mathcal{G}(\mathbf{x})|^2 - 2(\mathcal{G}(\mathbf{x}), \mathcal{G}(\mathbf{y})) + |\mathcal{G}(\mathbf{y})|^2. \end{aligned}$$

Putting these expressions into equality (8.33), we find that

$$|\mathcal{G}(\mathbf{x})|^2 - 2(\mathcal{G}(\mathbf{x}), \mathcal{G}(\mathbf{y})) + |\mathcal{G}(\mathbf{y})|^2 = |\mathbf{x}|^2 - 2(\mathbf{x}, \mathbf{y}) + |\mathbf{y}|^2. \quad (8.34)$$

Setting the vector  $\mathbf{y}$  equal to  $\mathbf{0}$  in relationship (8.34), and taking into account that  $\mathcal{G}(\mathbf{0}) = \mathbf{0}$ , we obtain the equality  $|\mathcal{G}(\mathbf{x})| = |\mathbf{x}|$  for all  $\mathbf{x} \in \mathbf{L}$ . Finally, taking into account the relationships  $|\mathcal{G}(\mathbf{x})| = |\mathbf{x}|$  and  $|\mathcal{G}(\mathbf{y})| = |\mathbf{y}|$ , from (8.34) follows the required equality (8.31).

Thus for any orthonormal basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , the vectors  $\mathbf{e}'_1, \dots, \mathbf{e}'_n$ , defined by the relationships  $\mathcal{G}(\mathbf{e}_i) = \mathbf{e}'_i$ , also form an orthonormal basis, in which the coordinates of the vector  $\mathbf{x} = x_1\mathbf{e}_1 + \dots + x_n\mathbf{e}_n$  are given by the formula  $x_i = (\mathbf{x}, \mathbf{e}_i)$ . From this we obtain that  $(\mathcal{G}(\mathbf{x}), \mathbf{e}'_i) = x_i$ , and this implies that

$$\mathcal{G}(\mathbf{x}) = x_1\mathbf{e}'_1 + \dots + x_n\mathbf{e}'_n,$$

that is, the constructed mapping  $\mathcal{G} : \mathbf{L} \rightarrow \mathbf{L}$  satisfies the condition of the lemma. From this it follows that  $\mathcal{G}$  is a linear transformation of the space  $\mathbf{L}$ , and by property (8.31), it is an orthogonal transformation.  $\square$

Let us note that along the way, we have proved the possibility of expressing an arbitrary motion  $f$  in the form of the product

$$f = \mathcal{T}_a g, \quad (8.35)$$

where  $\mathcal{T}_a$  is a translation, and  $g$  has a fixed point  $O$  and corresponds to some orthogonal transformation  $\mathcal{G}$  of the space  $\mathbf{L}$  (see Example 8.36). From the representation (8.35) and results of Sect. 8.3, it follows that two orthonormal frames of reference can be mapped into each other by a motion, and moreover, it is unique.



Now we shall make use of the arbitrariness in the selection of  $O$  in the representation (8.35) of the motion  $f$ . By formula (8.21), for a change in the point  $O$ , the vector  $\mathbf{a}$  in (8.35) is replaced by the vector  $\mathbf{a} + \mathcal{G}(\mathbf{c}) - \mathbf{c}$ , where for  $\mathbf{c}$ , one can take an arbitrary vector of the space  $L$ . We have the representation

$$\mathbf{c} = \mathbf{c}_0 + \mathbf{c}_1 + \cdots + \mathbf{c}_k, \quad \mathbf{c}_i \in L_i, \quad (8.40)$$

in the case of the decomposition (8.38), or else we have

$$\mathbf{c} = \mathbf{c}_0 + \mathbf{c}_1 + \cdots + \mathbf{c}_k + \mathbf{c}_{k+1}, \quad \mathbf{c}_i \in L_i, \quad (8.41)$$

in the case of the decomposition (8.39).

Since  $\mathcal{G}(\mathbf{x}) = \mathbf{x}$  for every vector  $\mathbf{x} \in L_0$ , the term  $\mathbf{c}_0$  makes no contribution to the vector  $\mathcal{G}(\mathbf{c}) - \mathbf{c}$  added to  $\mathbf{a}$ . For  $i > 0$ , the situation is precisely the reverse: the transformation  $\mathcal{G} - \mathcal{E}$  defines a *nonsingular* transformation in  $L_i$ . This follows from the fact that the kernel of the transformation  $\mathcal{G} - \mathcal{E}$  is equal to  $\{\mathbf{0}\}$ , which is obvious for a rotation through the angle  $\varphi_i$ ,  $0 < \varphi_i < 2\pi$ , in the plane and for the transformation  $-\mathcal{E}$  on a line. Therefore, the image of the transformation  $\mathcal{G} - \mathcal{E}$  in  $L_i$  is equal to the entire subspace  $L_i$  for  $i > 0$ . That is, every vector  $\mathbf{a}_i \in L_i$  can be represented in the form  $\mathbf{a}_i = \mathcal{G}(\mathbf{c}_i) - \mathbf{c}_i$ , where  $\mathbf{c}_i$  is some other vector of the same space  $L_i$ ,  $i > 0$ .

Thus in accordance with the representations (8.40) and (8.41), the vector  $\mathbf{a}$  can be written in the form  $\mathbf{a} = \mathbf{a}_0 + \mathbf{a}_1 + \cdots + \mathbf{a}_k$  or  $\mathbf{a} = \mathbf{a}_0 + \mathbf{a}_1 + \cdots + \mathbf{a}_k + \mathbf{a}_{k+1}$ , depending on whether the transformation  $\mathcal{G}$  is proper or improper. We may set  $\mathbf{a}_i = \mathcal{G}(\mathbf{c}_i) - \mathbf{c}_i$ , where the vectors  $\mathbf{c}_i$  are defined respectively by relationship (8.40) or (8.41). As a result, we obtain the equality

$$\mathbf{a} + \mathcal{G}(\mathbf{c}) - \mathbf{c} = \mathbf{a}_0,$$

meaning that by our selection of the point  $O$ , we can obtain that the vector  $\mathbf{a}$  is contained in the subspace  $L_0$ .

We have thus proved the following theorem.

**Theorem 8.39** *Every motion  $f$  of an affine Euclidean space  $V$  can be represented in the form*

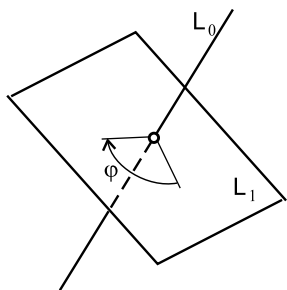
$$f = \mathcal{T}_a g, \quad (8.42)$$

where the transformation  $g$  has fixed point  $O$  and corresponds to the orthogonal transformation  $\mathcal{G} = \Lambda(g)$ , while  $\mathcal{T}_a$  is a translation by the vector  $\mathbf{a}$  such that  $\mathcal{G}(\mathbf{a}) = \mathbf{a}$ .

Let us consider the most visual example, that of the “physical” three-dimensional space in which we live. Here there are two possible cases.

*Case 1:* The motion  $f$  is proper. Then the orthogonal transformation  $\mathcal{G} : L \rightarrow L$  is also proper. Since  $\dim L = 3$ , the decomposition (8.38) has the form

$$L = L_0 \oplus L_1, \quad L_i \perp L_j,$$

**Fig. 8.5** A proper motion

where  $\dim L_0 = 1$  and  $\dim L_1 = 2$ . The transformation  $\mathcal{G}$  leaves vectors in  $L_0$  fixed and defines a rotation through the angle  $0 < \varphi < 2\pi$  in the plane  $L_1$ . Representation (8.42) shows that the transformation  $f$  can be obtained as a rotation through the angle  $\varphi$  about the line  $L_0$  and a translation in the direction of  $L_0$ ; see Fig. 8.5.

This result can be given a different formulation. Suppose a solid body executes an arbitrarily complex motion over time. Then its initial position can be superimposed on its final position by a rotation around some axis and a translation along that axis. Indeed, since it is a *solid* body, its final position is obtained from the initial position by some motion  $f$ . Since this change in position is obtained as a *continuous* motion, it follows that it is proper. Thus we may employ the three-dimensional case of Theorem 8.39. This result is known as *Euler's theorem*.

*Case 2:* The motion  $f$  is improper. Then the orthogonal transformation  $\mathcal{G} : L \rightarrow L$  is also improper. Since  $\dim L = 3$ , the decomposition (8.39) has the form

$$L = L_0 \oplus L_1 \oplus L_2, \quad L_i \perp L_j,$$

where  $L_0 = \{0\}$ ,  $\dim L_1 = 2$ , and  $\dim L_2 = 1$ . The transformation  $\mathcal{G}$  defines a rotation through the angle  $0 < \varphi < 2\pi$  in the plane  $L_1$  and carries each vector on the line  $L_2$  into its opposite. From this it follows that the equality  $\mathcal{G}(\mathbf{a}) = \mathbf{a}$  holds only for the vector  $\mathbf{a} = 0$ , and therefore, the translation  $\mathcal{T}_{\mathbf{a}}$  in formula (8.42) is equal to the identity transformation. Therefore, the motion  $f$  always has the fixed point  $O$ , and can be obtained as a rotation through the angle  $0 < \varphi < 2\pi$  in the plane  $L_1$  passing through this point followed by a reflection in the plane  $L_1$ .

The theory of motions in an affine Euclidean space can be given a more graphical form if we employ the notion of flags, which was introduced in Sect. 8.2 (p. 300). First, it is clear that a motion of a space carries a flag to a flag. The main result, which we in fact have already proved, can be formulated as follows.

**Theorem 8.40** *For every pair of flags, there exists a motion taking the first flag to the second, and such a motion is unique.*

*Proof* To prove the theorem, we observe that for an arbitrary flag

$$V_0 \subset V_1 \subset \cdots \subset V_n = V, \quad (8.43)$$

the affine subspace  $V_0$  consists by definition of a single point. Setting  $V_0 = O$ , we may identify each subspace  $V_i$  with the subspace  $L_i \subset L$ , where  $L_i$  is the space of vectors of the affine space  $V_i$ . Here the sequence

$$L_0 \subset L_1 \subset \cdots \subset L_n = L \quad (8.44)$$

defines a flag in  $L$ . On the other hand, we saw in Sect. 7.2 that the flag (8.44) is uniquely associated with an orthonormal basis  $e_1, \dots, e_n$  in  $L$ . Thus  $L_i = \langle e_1, \dots, e_i \rangle$  and  $e_i \in L_i^+$ , as established in Sect. 7.2. This means that the flag (8.43) is uniquely determined by some orthonormal frame of reference  $(O; e_1, \dots, e_n)$  in  $V$ . As we noted above, for two orthonormal frames of reference, there exists a unique motion of the space  $V$  taking the first frame of reference to the second. This holds, then, for two flags of the form (8.43), which proves the assertion of the theorem.  $\square$

The property proved in Theorem 8.40 is called “free mobility” of an affine Euclidean space. In the case of three-dimensional space, this assertion is a mathematical expression of the fact that in space, a solid body can be arbitrarily translated and rotated.

In an affine Euclidean space, the distance  $r(A, B)$  between any two points does not change under a motion of the space. In a general affine space it is impossible to associate with each pair of points a number that would be invariant under every non-singular affine transformation. This follows from the fact that for an arbitrary pair of points  $A, B$  and another arbitrary pair  $A', B'$ , there exists an affine transformation  $f$  taking  $A$  to  $A'$  and  $B$  to  $B'$ .

To prove this, let us write down a transformation  $f$  according to formula (8.19) in the form  $f = \mathcal{T}_a f_0$ , choosing the point  $A$  as the point  $O$ . Here  $A$  is a fixed point of the affine transformation  $f_0$ , that is,  $f_0(A) = A$ . The transformation  $f_0$  is defined by some linear transformation of the space of vectors  $L$  of our affine space  $V$  and is uniquely defined by the relation

$$\overrightarrow{Af_0(C)} = \mathcal{F}(\overrightarrow{AC}), \quad C \in V.$$

Then the condition  $f(A) = A'$  will be satisfied if we set  $a = \overrightarrow{AA'}$ . It remains to select a linear transformation  $\mathcal{F} : L \rightarrow L$  so as to satisfy the equality  $f(B) = B'$ , that is,  $\mathcal{T}_a f_0(B) = B'$ , which is equivalent to the relationship

$$f_0(B) = \mathcal{T}_{-a}(B'). \quad (8.45)$$

We set the vector  $x$  equal to  $\overrightarrow{AB}$  (under the condition  $A \neq B$ , whence  $x \neq 0$ ) and consider the point  $P = \mathcal{T}_{-a}(B')$  and vector  $y = \overrightarrow{AP}$ . Then the relationship (8.45) is equivalent to the equality  $\mathcal{F}(x) = y$ . It remains only to find a linear transformation  $\mathcal{F} : L \rightarrow L$  for which the condition  $\mathcal{F}(x) = y$  is satisfied for given vectors  $x$  and  $y$ , with  $x \neq 0$ . For this, we must extend the vector  $x$  to a basis of the space  $L$  and define  $\mathcal{F}$  in terms of the vectors of this basis arbitrarily, provided only that the condition  $\mathcal{F}(x) = y$  is satisfied.

# Chapter 9

## Projective Spaces

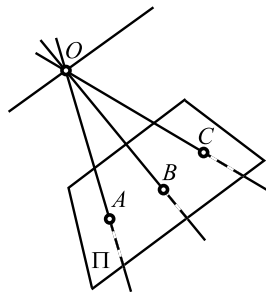
### 9.1 Definition of a Projective Space

In plane geometry, points and lines in the plane play very similar roles. In order to emphasize this symmetry, the fundamental property that connects points and lines in the plane is called *incidence*, and the fact that a point  $A$  lies on a line  $l$  or that a line  $l$  passes through a point  $A$  expresses in a symmetric form that  $A$  and  $l$  are *incident*. Then one might hope that to each assertion of geometry about incidence of points and lines there would correspond another assertion obtained from the first by everywhere interchanging the words “point” and “line.” And such is indeed the case, with some exceptions. For example, to every pair of distinct points, there is incident one and only one line. But it is not true that to every pair of distinct lines, there is incident one and only one point: the exception is the case that the lines are parallel. Then not a single point is incident to the two lines.

*Projective geometry* gives us the possibility of eliminating such exceptions by adding to the plane certain points called *points at infinity*. For example, if we do this, then two parallel lines will be incident at some point at infinity. And indeed, with a naive perception of the external world, we “see” that parallel lines moving away from us converge and intersect at a point on the “horizon.” Strictly speaking, the “horizon” is the totality of all points at infinity by which we extend the plane.

In analyzing this example, we may say that a point  $p$  of the plane seen by us corresponds to the point where the line passing through  $p$  and the center of our eye meets the retina. Mathematically, this situation is described using the notion of *central projection*.

Let us assume that the plane  $\Pi$  that we are investigating is contained in three-dimensional space. Let us choose in this same space some point  $O$  not contained in the plane  $\Pi$ . Every point  $A$  of the plane  $\Pi$  can be joined to  $O$  by the line  $OA$ . Conversely, a line passing through the point  $O$  intersects the plane  $\Pi$  in a certain point, *provided that the line is not parallel to  $\Pi$* . Thus most straight lines passing through the point  $O$  correspond to points  $A \in \Pi$ . But lines parallel to  $\Pi$  intuitively correspond precisely to *points at infinity* of the plane  $\Pi$ , or “points on the horizon.” See Fig. 9.1.

**Fig. 9.1** Central projection

We shall make this notion the basis of the definition of *projective space* and shall develop it in more detail in the sequel.

**Definition 9.1** Let  $L$  be a vector space of finite dimension. The collection of all lines  $\langle x \rangle$ , where  $x$  is a nonnull vector of the space  $L$ , is called a *projectivization* of  $L$  or *projective space*  $\mathbb{P}(L)$ . Here the lines  $\langle x \rangle$  themselves are called *points* of the projective space  $\mathbb{P}(L)$ . The *dimension* of the space  $\mathbb{P}(L)$  is defined as the number  $\dim \mathbb{P}(L) = \dim L - 1$ .

As we saw in Chap. 3, all vector spaces of a given dimension  $n$  are isomorphic. This fact is expressed by saying that there exists only one theory of  $n$ -dimensional vector spaces. In the same sense, there exists only one theory of  $n$ -dimensional projective space.

We shall frequently denote the projective space of dimension  $n$  by  $\mathbb{P}^n$  if we have no need of indicating the  $(n + 1)$ -dimensional vector space on the basis of which it was constructed.

If  $\dim \mathbb{P}(L) = 1$ , then  $\mathbb{P}(L)$  is called the *projective line*, and if  $\dim \mathbb{P}(L) = 2$ , then it is called the *projective plane*. Lines in an ordinary plane are points on the projective line, while lines in three-dimensional space are points in the projective plane.

And as earlier, we give the reader the choice whether to consider  $L$  a real or complex space, or even to consider it as a space over an arbitrary field  $\mathbb{K}$  (with the exception of certain questions related specifically to real spaces). In accordance with the definition given above, we shall say that  $\dim \mathbb{P}(L) = -1$  if  $\dim L = 0$ . In this case, the set  $\mathbb{P}(L)$  is empty.

In order to introduce coordinates in a space  $\mathbb{P}(L)$  of dimension  $n$ , we choose a basis  $e_0, e_1, \dots, e_n$  in the space  $L$ . A point  $A \in \mathbb{P}(L)$  is by definition a line  $\langle x \rangle$ , where  $x$  is some nonnull vector in  $L$ . Thus we have the representation

$$x = \alpha_0 e_0 + \alpha_1 e_1 + \dots + \alpha_n e_n. \quad (9.1)$$

The numbers  $(\alpha_0, \alpha_1, \dots, \alpha_n)$  are called *homogeneous coordinates* of the point  $A$ . But the point  $A$  is the entire line  $\langle x \rangle$ . It can also be obtained in the form  $\langle y \rangle$  if  $y = \lambda x$  and  $\lambda \neq 0$ . Then

$$y = \lambda \alpha_0 e_0 + \lambda \alpha_1 e_1 + \dots + \lambda \alpha_n e_n.$$

From this it follows that the numbers  $(\lambda\alpha_0, \lambda\alpha_1, \dots, \lambda\alpha_n)$  are also homogeneous coordinates of the point  $A$ . That is, homogeneous coordinates are defined only up to a common nonzero factor. Since by definition,  $A = \langle \mathbf{x} \rangle$  and  $\mathbf{x} \neq \mathbf{0}$ , they cannot all be simultaneously equal to zero. In order to emphasize that homogeneous coordinates are defined only up to a nonzero common factor, they are written in the form

$$(\alpha_0 : \alpha_1 : \alpha_2 : \dots : \alpha_n). \quad (9.2)$$

Thus if we wish to express some property of the point  $A$  in terms of its homogeneous coordinates, then that assertion must continue to hold if all the homogeneous coordinates  $(\alpha_0, \alpha_1, \dots, \alpha_n)$  are simultaneously multiplied by the same nonzero number.

Let us assume, for example, that we are considering the points of projective space whose homogeneous coordinates satisfy the relationship

$$F(\alpha_0, \alpha_1, \dots, \alpha_n) = 0, \quad (9.3)$$

where  $F$  is a polynomial in  $n + 1$  variables. In order for this requirement actually to be related to the points and not depend on the factor  $\lambda$  by which we can multiply their homogeneous coordinates, it is necessary that along with the numbers  $(\alpha_0, \alpha_1, \dots, \alpha_n)$ , the relationship (9.3) be satisfied as well by the numbers  $(\lambda\alpha_0, \lambda\alpha_1, \dots, \lambda\alpha_n)$  for an arbitrary nonzero factor  $\lambda$ .

Let us elucidate when this requirement is satisfied. To this end, in the polynomial  $F(x_0, x_1, \dots, x_n)$  let us collect all terms of the form  $a x_0^{k_0} x_1^{k_1} \dots x_n^{k_n}$  with  $k_0 + k_1 + \dots + k_n = m$  and denote their sum by  $F_m$ . We thereby obtain the representation

$$F(x_0, x_1, \dots, x_n) = \sum_{m=0}^N F_m(x_0, x_1, \dots, x_n).$$

It follows at once from the definition of  $F_m$  that

$$F_m(\lambda x_0, \lambda x_1, \dots, \lambda x_n) = \lambda^m F_m(x_0, x_1, \dots, x_n).$$

From this, we obtain

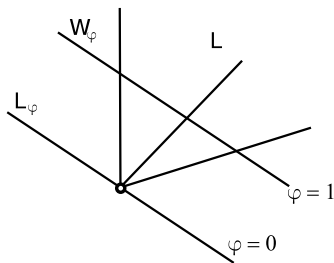
$$F(\lambda x_0, \lambda x_1, \dots, \lambda x_n) = \sum_{m=0}^N \lambda^m F_m(x_0, x_1, \dots, x_n).$$

Our condition means that the equality  $\sum_{m=0}^N \lambda^m F_m = 0$  is satisfied for the coordinates of the points in question and simultaneously for all nonzero values of  $\lambda$ . Let us denote by  $c_m$  the value  $F_m(\alpha_0, \alpha_1, \dots, \alpha_n)$  for some concrete choice of homogeneous coordinates  $(\alpha_0, \alpha_1, \dots, \alpha_n)$ . Then we arrive at the condition  $\sum_{m=0}^N c_m \lambda^m = 0$  for all nonzero values  $\lambda$ . This means that the polynomial  $\sum_{m=0}^N c_m \lambda^m$  in the variable  $\lambda$  has an infinite number of roots (for simplicity, we are now assuming that the field  $\mathbb{K}$  over which the vector space  $L$  is being considered is infinite; however, it would be possible to eliminate this restriction). Then, by a well-known theorem on





**Fig. 9.2** Affine subset of a projective space



$i$ th homogeneous coordinate is nonzero by  $V_i$ . It is obvious that  $V_i$  is already not a projective subspace in  $\mathbb{P}(L)$ .

The following construction is a natural generalization of Example 9.4. In the space  $L$  let an arbitrary basis  $e_0, e_1, \dots, e_n$  be chosen. Let us consider some linear function  $\varphi$  on the space  $L$  not identically equal to zero. Vectors  $x \in L$  for which  $\varphi(x) = 0$  form a hyperplane  $L_\varphi \subset L$ . It is a subspace of the solutions of the “system” consisting of a single linear homogeneous equation. To it is associated the projective hyperplane  $\mathbb{P}(L_\varphi) \subset \mathbb{P}(L)$ . It is obvious that  $L_\varphi$  coincides with the hyperplane  $L_i$  from Example 9.4 if the linear function  $\varphi$  maps each vector  $x \in L$  onto its  $i$ th coordinate, that is,  $\varphi$  is the  $i$ th vector of the basis of the space  $L^*$ , the dual of the basis  $e_0, e_1, \dots, e_n$  of the space  $L$ .

Let us now denote by  $W_\varphi$  the set of vectors  $x \in L$  for which  $\varphi(x) = 1$ . This is again the set of solutions of the “system” consisting of a single linear equation, but now inhomogeneous. It can be viewed naturally as an affine space with space of vectors  $L_\varphi$ . Let us denote the set  $\mathbb{P}(L) \setminus \mathbb{P}(L_\varphi)$  by  $V_\varphi$ . Then for every point  $A \in V_\varphi$  there exists a unique vector  $x \in W_\varphi$  for which  $A = \langle x \rangle$ .

In this way, we may identify the set  $V_\varphi$  with the set  $W_\varphi$ , and with the aid of this identification, consider  $V_\varphi$  an affine space. By definition, its space of vectors is  $L_\varphi$ , and if  $A$  and  $B$  are two points in  $V_\varphi$ , then there exist two vectors  $x$  and  $y$  for which  $\varphi(x) = 1$  and  $\varphi(y) = 1$  such that  $A = \langle x \rangle$  and  $B = \langle y \rangle$ , and then  $\overrightarrow{AB} = y - x$ . Thus the  $n$ -dimensional projective space  $\mathbb{P}(L)$  can be represented as the union of the  $n$ -dimensional affine space  $V_\varphi$  and the projective hyperplane  $\mathbb{P}(L_\varphi) \subset \mathbb{P}(L)$ ; see Fig. 9.2. In the sequel, we shall call  $V_\varphi$  an *affine subset* of the space  $\mathbb{P}(L)$ .

Let us choose in the space  $L$  a basis  $e_0, \dots, e_n$  such that  $\varphi(e_0) = 1$  and  $\varphi(e_i) = 0$  for all  $i = 1, \dots, n$ . Then the vector  $e_0$  is associated with the point  $O = \langle e_0 \rangle$  belonging to the affine subset  $V_\varphi$ , while all the remaining vectors  $e_1, \dots, e_n$  are in  $L_\varphi$ , and they are associated with the points  $\langle e_1 \rangle, \dots, \langle e_n \rangle$  lying in the hyperplane  $\mathbb{P}(L_\varphi)$ . We have thus constructed in the affine space  $(V_\varphi, L_\varphi)$  a frame of reference  $(O; e_1, \dots, e_n)$ . The coordinates  $(\xi_1, \dots, \xi_n)$  of the point  $A \in V_\varphi$  with respect to this frame of reference are called *inhomogeneous coordinates* of the point  $A$  in our projective space. We wish to emphasize that they are defined only for points in the affine subset  $V_\varphi$ . If we return to the definitions, then we see that the inhomogeneous coordinates  $(\xi_1, \dots, \xi_n)$  are obtained from the homogeneous coordinates

(9.2) through the formula

$$\xi_i = \frac{\alpha_i}{\alpha_0}, \quad i = 1, \dots, n. \quad (9.6)$$

It is obvious here that for  $\mathbf{x}$  from formula (9.1), the function  $\varphi$  that we have chosen assumes the value  $\varphi(\mathbf{x}) = \alpha_0$ .

In order to extend the concept of inhomogeneous coordinates to all points of a projective space  $\mathbb{P}(L) = V_\varphi \cup \mathbb{P}(L_\varphi)$ , it remains also to consider the points of the projective hyperplane  $\mathbb{P}(L_\varphi)$ . For such points it is natural to assign the value  $\alpha_0 = 0$ . Sometimes this is expressed by saying that the inhomogeneous coordinates  $(\xi_1, \dots, \xi_n)$  of the point  $A \in \mathbb{P}(L_\varphi)$  assume *infinite values*, which justifies thinking of  $\mathbb{P}(L_\varphi)$  as a set of “points at infinity” (horizon) for the affine subset  $V_\varphi$ .

Of course, one could also choose a linear function  $\varphi$  such that  $\varphi(\mathbf{e}_i) = 1$  for some number  $i \in \{0, \dots, n\}$ , not necessarily equal to 0, as was done above, and  $\varphi(\mathbf{e}_j) = 0$  for all  $j \neq i$ . We will denote the associated spaces  $V_\varphi$  and  $L_\varphi$  by  $V_i$  and  $L_i$ . In this case, the projective space  $\mathbb{P}(L)$  can be represented in the analogous form  $V_i \cup \mathbb{P}(L_i)$ , that is, as the union of an affine part  $V_i$  and a hyperplane  $\mathbb{P}(L_i)$  for the corresponding value  $i \in \{0, \dots, n\}$ . Sometimes this fact is expressed by saying that in the projective space  $\mathbb{P}(L)$ , one may introduce various *affine charts*. It is not difficult to see that every point  $A$  of a projective space  $\mathbb{P}(L)$  is “finite” for some value  $i \in \{0, \dots, n\}$ , that is, it belongs to the subset  $V_i$  for the corresponding value  $i$ . This follows from the fact that by definition, homogeneous coordinates (9.2) of the point  $A$  are not simultaneously equal to zero. If  $\alpha_i \neq 0$  for some  $i \in \{0, \dots, n\}$ , then  $A$  is contained in the associated affine subset  $V_i$ .

If  $L'$  and  $L''$  are two subspaces of a space  $L$ , then it is obvious that

$$\mathbb{P}(L') \cap \mathbb{P}(L'') = \mathbb{P}(L' \cap L''). \quad (9.7)$$

It is somewhat more complicated to interpret the set  $\mathbb{P}(L' + L'')$ . It is obvious that it does not coincide with  $\mathbb{P}(L') \cup \mathbb{P}(L'')$ . For example, if  $L'$  and  $L''$  are two distinct lines in the plane  $L$ , then the set  $\mathbb{P}(L') \cup \mathbb{P}(L'')$  consisting of two points is in general not a projective subspace of the space  $\mathbb{P}(L)$ .

To give a geometric interpretation to the sets  $\mathbb{P}(L' + L'')$ , we shall introduce the following notion. Let  $P = \langle \mathbf{e} \rangle$  and  $P' = \langle \mathbf{e}' \rangle$  be two distinct points of the projective space  $\mathbb{P}(L)$ . Let us set  $L_1 = \langle \mathbf{e}, \mathbf{e}' \rangle$  and consider the one-dimensional projective subspace  $\mathbb{P}(L_1)$ . It obviously contains both points  $P$  and  $P'$ , and moreover, it is contained in every projective subspace containing the points  $P$  and  $P'$ . Indeed, if  $L_2 \subset L$  is a vector subspace such that  $\mathbb{P}(L_2)$  contains the points  $P$  and  $P'$ , then this means that  $L_2$  contains the vectors  $\mathbf{e}$  and  $\mathbf{e}'$ , which implies that it also contains the entire subspace  $L_1 = \langle \mathbf{e}, \mathbf{e}' \rangle$ . Therefore, by the definition of a projective subspace, we have that  $\mathbb{P}(L_1) \subset \mathbb{P}(L_2)$ .

**Definition 9.5** The one-dimensional projective subspace  $\mathbb{P}(L_1)$  constructed from two given points  $P \neq P'$  is called the *line* connecting the points  $P$  and  $P'$ .

**Theorem 9.6** *Let  $L'$  and  $L''$  be two subspaces of a vector space  $L$ . Then the union of lines connecting all possible points of  $\mathbb{P}(L')$  with all possible points of  $\mathbb{P}(L'')$  coincides with the projective subspace  $\mathbb{P}(L' + L'')$ .*

*Proof* We shall denote by  $\Sigma$  the union of lines described in the statement of the theorem. Every such line has the form  $\mathbb{P}(L_1)$ , where  $L_1 = \langle e', e'' \rangle$ , for vectors  $e' \in L'$  and  $e'' \in L''$ . Since  $e' + e'' \in L' + L''$ , it follows from the preceding discussion that every such line  $\mathbb{P}(L_1)$  belongs to  $\mathbb{P}(L' + L'')$ . Thus we have proved the set inclusion  $\Sigma \subset \mathbb{P}(L' + L'')$ .

Conversely, suppose now that the point  $S \in \mathbb{P}(L)$  belongs to the projective subspace  $\mathbb{P}(L' + L'')$ . This means that  $S = \langle e \rangle$ , where the vector  $e$  is in  $L' + L''$ . And this implies that the vector  $e$  can be represented in the form  $e = e' + e''$ , where  $e' \in L'$  and  $e'' \in L''$ . This means that  $S = \langle e \rangle$  and the vector  $e$  belongs to the plane  $\langle e', e'' \rangle$ , that is,  $S$  lies on the line connecting the point  $\langle e' \rangle$  in  $\mathbb{P}(L')$  to the point  $\langle e'' \rangle$  in  $\mathbb{P}(L'')$ . In other words, we have  $S \in \Sigma$ , and thus the subspace  $\mathbb{P}(L' + L'')$  is contained in  $\Sigma$ . Taking into account the reverse inclusion proved above, we obtain the required equality  $\Sigma = \mathbb{P}(L' + L'')$ .  $\square$

**Definition 9.7** The set  $\mathbb{P}(L' + L'')$  is called a *projective cover* of the set  $\mathbb{P}(L') \cup \mathbb{P}(L'')$  and is denoted by

$$\mathbb{P}(L' + L'') = \overline{\mathbb{P}(L') \cup \mathbb{P}(L'')}. \quad (9.8)$$

Recalling Theorem 3.41, we obtain the following result.

**Theorem 9.8** *If  $\mathbb{P}'$  and  $\mathbb{P}''$  are two projective subspaces of a projective space  $\mathbb{P}(L)$ , then*

$$\dim(\mathbb{P}' \cap \mathbb{P}'') + \dim(\overline{\mathbb{P}' \cup \mathbb{P}''}) = \dim \mathbb{P}' + \dim \mathbb{P}''. \quad (9.9)$$

**Example 9.9** If  $\mathbb{P}'$  and  $\mathbb{P}''$  are two lines in the projective plane  $\mathbb{P}(L)$ ,  $\dim L = 3$ , then  $\dim \mathbb{P}' = \dim \mathbb{P}'' = 1$  and  $\dim(\overline{\mathbb{P}' \cup \mathbb{P}''}) \leq 2$ , and from relationship (9.9), we obtain that  $\dim(\mathbb{P}' \cap \mathbb{P}'') \geq 0$ , that is, every pair of lines in the projective plane intersect.

The theory of projective spaces exhibits a beautiful symmetry, which goes under the name *duality* (we have already encountered an analogous phenomenon in the theory of vector spaces; see Sect. 3.7).

Let  $L^*$  be the dual space to  $L$ . The projective space  $\mathbb{P}(L^*)$  is called the *dual* of  $\mathbb{P}(L)$ . Every point of the dual space  $\mathbb{P}(L^*)$  is by definition a line  $\langle f \rangle$ , where  $f$  is a linear function on the space  $L$  not identically zero. Such a function determines a hyperplane  $L_f \subset L$ , given by the linear homogeneous equation  $f(x) = 0$  in the vector space  $L$ , which means that the hyperplane  $\mathbb{P}_f$  is equal to  $\mathbb{P}(L_f)$  in the projective space  $\mathbb{P}(L)$ .

Let us prove that the correspondence constructed above between points  $\langle f \rangle$  of the dual space  $\mathbb{P}(L^*)$  and hyperplanes  $\mathbb{P}_f$  of the space  $\mathbb{P}(L)$  is a bijection. To do so, we must prove that the equations  $f = 0$  and  $\alpha f = 0$  are equivalent, defining one and the

same hyperplane, that is,  $\mathbb{P}_f = \mathbb{P}_{\alpha f}$ . As was shown in Sect. 3.7, every hyperplane  $L' \subset L$  is determined by a single nonzero linear equation. Two different equations  $f = 0$  and  $f_1 = 0$  can define one and the same hyperplane only if  $f_1 = \alpha f$ , where  $\alpha$  is some nonzero number. Indeed, in the contrary case, the system of the two equations  $f = 0$  and  $f_1 = 0$  has rank 2, and therefore, it defines a subspace  $L''$  of dimension  $n - 2$  in  $L$  and a subspace  $\mathbb{P}(L'') \subset \mathbb{P}(L)$  of dimension  $n - 3$ , which is obviously not a hyperplane. Thus the dual space  $\mathbb{P}(L^*)$  can be interpreted as the space of hyperplanes in  $\mathbb{P}(L)$ . This is the simplest example of the fact that certain geometric objects cannot be described by numbers (such as, for example, vector spaces can be described by their dimension), but constitute a set having a geometric character. We shall encounter more complex examples in Chap. 10.

There is also a much more general fact, namely that there is a bijection between  $m$ -dimensional projective subspaces of the space  $\mathbb{P}(L)$  (dimension  $n$ ) and subspaces of dimension  $n - m - 1$  of the space  $\mathbb{P}(L^*)$ . We shall now describe this correspondence, and the reader will easily verify that for  $m = n - 1$ , this coincides with the above-described correspondence between hyperplanes in  $\mathbb{P}(L)$  and points in  $\mathbb{P}(L^*)$ .

Let  $L' \subset L$  be a subspace of dimension  $m + 1$ , so that  $\dim \mathbb{P}(L') = m$ . Let us consider in the dual space  $L^*$ , the annihilator  $(L')^a$  of the subspace  $L'$ . Let us recall that the annihilator is the subspace  $(L')^a \subset L^*$  consisting of all linear functions  $f \in L^*$  such that  $f(x) = 0$  for all vectors  $x \in L'$ . As we established in Sect. 3.7 (formula (3.54)), the dimension of the annihilator is equal to

$$\dim(L')^a = \dim L - \dim L' = n - m. \quad (9.10)$$

The projective subspace  $\mathbb{P}((L')^a) \subset \mathbb{P}(L^*)$  is called the *dual* to the subspace  $\mathbb{P}(L') \subset \mathbb{P}(L)$ . By (9.10), its dimension is  $n - m - 1$ . What we have here is a variant of a concept that is well known to us. If a nonsingular symmetric bilinear form  $(x, y)$  is defined on the space  $L$ , then we can identify  $(L')^a$  with the orthogonal complement to  $L'$ , which was denoted by  $(L')^\perp$ ; see p. 198. If we write the bilinear form  $(x, y)$  in some orthonormal basis of the space  $L$ , then it takes the form  $\sum_{i=0}^n x_i y_i$ , and the point with coordinates  $(y_0, y_1, \dots, y_n)$  will correspond to the hyperplane defined by the equation

$$\sum_{i=0}^n x_i y_i = 0,$$

in which  $y_0, \dots, y_n$  are taken as fixed, and  $x_0, \dots, x_n$  are variables.

The assertions we have proved together with the duality principle established in Sect. 3.7 leads automatically to the following result, called the *principle of projective duality*.

**Proposition 9.10** (Principle of projective duality) *If a theorem is proved for all projective spaces of a given finite dimension  $n$  over a given field  $\mathbb{K}$  in a formulation that uses only the concepts of projective subspace, dimension, projective cover, and intersection, then for all such spaces, one has also the dual theorem obtained from*

the original one by the following substitutions:

$$\begin{array}{l|l} \text{dimension } m & \text{dimension } n - m - 1 \\ \text{intersection } \mathbb{P}_1 \cap \mathbb{P}_2 & \text{projective cover } \overline{\mathbb{P}_1 \cup \mathbb{P}_2} \\ \text{projective cover } \overline{\mathbb{P}_1 \cup \mathbb{P}_2} & \text{intersection } \mathbb{P}_1 \cap \mathbb{P}_2. \end{array}$$

For example, the assertion “through two distinct points of the projective plane there passes one line” has as its dual assertion “every pair of distinct lines in the projective plane intersect in one point.”

One may try to extend this principle in such a way that it will cover not only projective spaces, but also the projective algebraic varieties described by equation (9.5). However, in this regard there appear some new difficulties, which we shall only mention here without going into detail.

Assume, for example, that a projective algebraic variety  $X \subset \mathbb{P}(L)$  is given by the single equation

$$F(x_0, x_1, \dots, x_n) = 0,$$

where  $F$  is a homogeneous polynomial. To every point  $A \in X$  there corresponds a hyperplane given by the equation

$$\sum_{i=0}^n \frac{\partial F}{\partial x_i}(A) x_i = 0, \quad (9.11)$$

called the *tangent hyperplane* to  $X$  at the point  $A$  (this notion will be discussed later in greater detail). By the above considerations, we can assign to this hyperplane the point  $B$  of the dual space  $\mathbb{P}(L^*)$ .

It is natural to suppose that as  $A$  runs through all points  $X$ , then the point  $B$  also runs through some projective algebraic variety in the space  $\mathbb{P}(L)$ , called the *dual* to the original variety  $X$ . This is indeed the case, except for certain unpleasant exceptions. Namely, for some point  $A$ , it could be the case that all partial derivatives  $\frac{\partial F}{\partial x_i}(A)$  are equal to 0 for  $i = 0, 1, \dots, n$ , and equation (9.11) takes the form of the identity  $0 = 0$ . Such points are called *singular points* of the projective algebraic variety  $X$ . In this case, we do not obtain any hyperplane, and therefore, we cannot use the indicated method to assign to the point  $A$  a given point of the space  $\mathbb{P}(L^*)$ . It is possible to prove that singular points are in some sense exceptional. Moreover, many very interesting varieties have no singular points at all, so that for them, the dual variety exists. But then in the dual variety, there appear singular points, so that the beautiful symmetry nevertheless disappears. Overcoming all these difficulties is the task of *algebraic geometry*. We shall not go deeply into this, and we have mentioned it only in connection to the fact that in Chap. 11, devoted to quadrics, we shall consider precisely the special case in which these difficulties do not appear.

## 9.2 Projective Transformations

Let  $\mathcal{A}$  be a linear transformation of a vector space  $L$  into itself. It is natural to entertain the idea of extending it to the projective space  $\mathbb{P}(L)$ . It would seem to be something easy to do: one has only to associate with each point  $P \in \mathbb{P}(L)$  corresponding to the line  $\langle e \rangle$  in  $L$ , the line  $\langle \mathcal{A}(e) \rangle$ , which is some point of the projective space  $\mathbb{P}(L)$ . However, here we encounter the following difficulty: If  $\mathcal{A}(e) = \mathbf{0}$ , then we cannot construct the line  $\langle \mathcal{A}(e) \rangle$ , since all vectors proportional to  $\mathcal{A}(e)$  are the null vector. Thus the transformation that we wish to construct is not defined in general for all points of the projective space  $\mathbb{P}(L)$ . However, if we wished to define it for all points, then we must require that the kernel of the transformation  $\mathcal{A}$  be  $\{\mathbf{0}\}$ . As we know, this condition is equivalent to the transformation  $\mathcal{A} : L \rightarrow L$  being nonsingular. Thus to all nonsingular transformations  $\mathcal{A}$  of the space  $L$  into itself (and only these) there correspond mappings of the projective space  $\mathbb{P}(L)$  into itself. We shall denote them by  $\mathbb{P}(\mathcal{A})$ .

We have seen that a nonsingular transformation  $\mathcal{A} : L \rightarrow L$  defines a bijective mapping of the space  $L$  into itself. Let us prove that in this case, the corresponding mapping  $\mathbb{P}(\mathcal{A}) : \mathbb{P}(L) \rightarrow \mathbb{P}(L)$  is also a bijection. First, let us verify that its image coincides with all  $\mathbb{P}(L)$ . Let  $P$  be a point of the space  $\mathbb{P}(L)$ . It corresponds to some line  $\langle e \rangle$  in  $L$ . Since the transformation  $\mathcal{A}$  is nonsingular, it follows that  $e = \mathcal{A}(e')$  for some vector  $e' \in L$ , and moreover,  $e' \neq \mathbf{0}$ , since  $e \neq \mathbf{0}$ . If  $P'$  is a point of the space  $\mathbb{P}(L)$  corresponding to the line  $\langle e' \rangle$ , then  $P = \mathbb{P}(\mathcal{A})(P')$ . It remains to show that  $\mathbb{P}(\mathcal{A})$  cannot map two distinct points into one. Let us suppose that  $P \neq P'$  and

$$\mathbb{P}(\mathcal{A})(P) = \mathbb{P}(\mathcal{A})(P') = \overline{P}, \quad (9.12)$$

where the points  $P$ ,  $P'$ , and  $\overline{P}$  correspond to the lines  $\langle e \rangle$ ,  $\langle e' \rangle$ , and  $\langle \overline{e} \rangle$  respectively.

The condition  $P \neq P'$  is equivalent to the vectors  $e$  and  $e'$  being linearly independent, while from equality (9.12) it follows that  $\langle \mathcal{A}(e) \rangle = \langle \mathcal{A}(e') \rangle = \langle \overline{e} \rangle$ , which means that the vectors  $\mathcal{A}(e)$  and  $\mathcal{A}(e')$  are linearly dependent. But if  $\alpha \mathcal{A}(e) + \beta \mathcal{A}(e') = \mathbf{0}$ , where  $\alpha \neq 0$  or  $\beta \neq 0$ , then  $\mathcal{A}(\alpha e + \beta e') = \mathbf{0}$ , and since the transformation  $\mathcal{A}$  is nonsingular, we have  $\alpha e + \beta e' \neq \mathbf{0}$ , which contradicts the condition  $P \neq P'$ . Thus we have proved that the mapping  $\mathbb{P}(\mathcal{A}) : \mathbb{P}(L) \rightarrow \mathbb{P}(L)$  is a bijection. Consequently, the inverse mapping  $\mathbb{P}(\mathcal{A})^{-1}$  is also defined.

**Definition 9.11** A mapping  $\mathbb{P}(\mathcal{A})$  of the projective space  $\mathbb{P}(L)$  corresponding to the nonsingular transformation  $\mathcal{A}$  of a vector space  $L$  into itself is called a *projective transformation* of the space  $\mathbb{P}(L)$ .

**Theorem 9.12** We have the following assertions:

- (1)  $\mathbb{P}(\mathcal{A}_1) = \mathbb{P}(\mathcal{A}_2)$  if and only if  $\mathcal{A}_2 = \lambda \mathcal{A}_1$ , where  $\lambda$  is some nonzero scalar.
- (2) If  $\mathcal{A}_1$  and  $\mathcal{A}_2$  are two nonsingular transformations of a vector space  $L$ , then  $\mathbb{P}(\mathcal{A}_1 \mathcal{A}_2) = \mathbb{P}(\mathcal{A}_1) \mathbb{P}(\mathcal{A}_2)$ .
- (3) If  $\mathcal{A}$  is a nonsingular transformation, then  $\mathbb{P}(\mathcal{A})^{-1} = \mathbb{P}(\mathcal{A}^{-1})$ .

(4) A projective transformation  $\mathbb{P}(\mathcal{A})$  carries every projective subspace of the space  $\mathbb{P}(\mathbf{L})$  into a subspace of the same dimension.

*Proof* All the assertions of the proof follow directly from the definitions.

(1) If  $\mathcal{A}_2 = \lambda \mathcal{A}_1$ , then it is obvious that  $\mathcal{A}_1$  and  $\mathcal{A}_2$  map lines of the vector space  $\mathbf{L}$  in exactly the same way, that is,  $\mathbb{P}(\mathcal{A}_1) = \mathbb{P}(\mathcal{A}_2)$ . Now suppose, conversely, that  $\mathbb{P}(\mathcal{A}_1)(A) = \mathbb{P}(\mathcal{A}_2)(A)$  for an arbitrary point  $A \in \mathbb{P}(\mathbf{L})$ . If the point  $A$  corresponds to the line  $\langle \mathbf{e} \rangle$ , then we have  $\langle \mathcal{A}_1(\mathbf{e}) \rangle = \langle \mathcal{A}_2(\mathbf{e}) \rangle$ , that is,

$$\mathcal{A}_2(\mathbf{e}) = \lambda \mathcal{A}_1(\mathbf{e}), \quad (9.13)$$

where  $\lambda$  is some scalar. However, in theory, the number  $\lambda$  in relationship (9.13) could have had its own value for each vector  $\mathbf{e}$ . Let us consider two linearly independent vectors  $\mathbf{x}$  and  $\mathbf{y}$  and for the vectors  $\mathbf{x}$ ,  $\mathbf{y}$ , and  $\mathbf{x} + \mathbf{y}$ , let us write down condition (9.13):

$$\begin{cases} \mathcal{A}_2(\mathbf{x}) = \lambda \mathcal{A}_1(\mathbf{x}), \\ \mathcal{A}_2(\mathbf{y}) = \mu \mathcal{A}_1(\mathbf{y}), \\ \mathcal{A}_2(\mathbf{x} + \mathbf{y}) = \nu \mathcal{A}_1(\mathbf{x} + \mathbf{y}). \end{cases} \quad (9.14)$$

In view of the linearity of  $\mathcal{A}_1$  and  $\mathcal{A}_2$ , we have

$$\mathcal{A}_1(\mathbf{x} + \mathbf{y}) = \mathcal{A}_1(\mathbf{x}) + \mathcal{A}_1(\mathbf{y}), \quad \mathcal{A}_2(\mathbf{x} + \mathbf{y}) = \mathcal{A}_2(\mathbf{x}) + \mathcal{A}_2(\mathbf{y}). \quad (9.15)$$

Having substituted expressions (9.15) into the third equality of (9.14), we then subtract from it the first and second inequalities. We then obtain

$$(\nu - \lambda)\mathcal{A}_1(\mathbf{x}) + (\nu - \mu)\mathcal{A}_1(\mathbf{y}) = \mathcal{A}_1((\nu - \lambda)\mathbf{x} + (\nu - \mu)\mathbf{y}) = \mathbf{0}.$$

Since the transformation  $\mathcal{A}_1$  is nonsingular (by the definition of a projective transformation), it follows that  $(\nu - \lambda)\mathbf{x} + (\nu - \mu)\mathbf{y} = \mathbf{0}$ , and in view of the linear independence of the vectors  $\mathbf{x}$  and  $\mathbf{y}$ , it follows from this that  $\lambda = \nu$  and  $\mu = \nu$ , that is, all the scalars  $\lambda, \mu, \nu$  in (9.14) are the same, and therefore the scalar  $\lambda$  in relationship (9.13) is one and the same for all vectors  $\mathbf{e} \in \mathbf{L}$ .

(2) We must prove that for every point  $P$  of the corresponding line  $\langle \mathbf{e} \rangle$ , we have the equality  $\mathbb{P}(\mathcal{A}_1 \mathcal{A}_2)(P) = \mathbb{P}(\mathcal{A}_1)(\mathbb{P}(\mathcal{A}_2)(P))$ , and this, by the definition of a projective transformation, follows from the fact that  $\langle (\mathcal{A}_1 \mathcal{A}_2)(\mathbf{e}) \rangle = \mathcal{A}_1(\langle \mathcal{A}_2(\mathbf{e}) \rangle)$ . The last equality follows from the definition of the product of linear transformations.

(3) By what we have proven, we have the equality  $\mathbb{P}(\mathcal{A})\mathbb{P}(\mathcal{A}^{-1}) = \mathbb{P}(\mathcal{A}\mathcal{A}^{-1}) = \mathbb{P}(\mathcal{E})$ . It is obvious that  $\mathbb{P}(\mathcal{E})$  is the identity transformation of the space  $\mathbb{P}(\mathbf{L})$  into itself. From this, it follows that  $\mathbb{P}(\mathcal{A})^{-1} = \mathbb{P}(\mathcal{A}^{-1})$ .

(4) Finally, let  $\mathbf{L}'$  be an  $m$ -dimensional subspace of the vector space  $\mathbf{L}$  and let  $\mathbb{P}(\mathbf{L}')$  be the associated  $(m - 1)$ -dimensional projective subspace. The mapping  $\mathbb{P}(\mathcal{A})$  takes  $\mathbb{P}(\mathbf{L}')$  into a collection of points of the form  $P'' = \langle \mathcal{A}(\mathbf{e}') \rangle$ , where  $P' = \langle \mathbf{e}' \rangle$  runs through all points of  $\mathbb{P}(\mathbf{L}')$ . This holds because  $\mathbf{e}'$  runs through all vectors of the space  $\mathbf{L}'$ . Let us prove that here, all vectors  $\langle \mathcal{A}(\mathbf{e}') \rangle$  coincide with



the nonnull vectors of some vector subspace  $L''$  having the same dimension as  $L'$ . This will give us the required assertion.

In the subspace  $L'$ , let us choose a basis  $e_1, \dots, e_m$ . Then every vector  $e' \in L'$  can be represented in the form

$$e' = \alpha_1 e_1 + \dots + \alpha_m e_m,$$

while the condition  $e' \neq 0$  is equivalent to not all the coefficients  $\alpha_i$  being equal to zero. From this, we obtain

$$\mathcal{A}(e') = \alpha_1 \mathcal{A}(e_1) + \dots + \alpha_m \mathcal{A}(e_m). \quad (9.16)$$

The vectors  $\mathcal{A}(e_1), \dots, \mathcal{A}(e_m)$  are linearly independent, since the transformation  $\mathcal{A} : L \rightarrow L$  is nonsingular. Let us consider the  $m$ -dimensional subspace  $L'' = \langle \mathcal{A}(e_1), \dots, \mathcal{A}(e_m) \rangle$ . From the relationship (9.16), it follows that the transformation  $\mathbb{P}(\mathcal{A})$  takes the points of the subspace  $\mathbb{P}(L')$  precisely into the points of the subspace  $\mathbb{P}(L'')$ . From the equality  $\dim L' = \dim L'' = m$ , we obtain  $\dim \mathbb{P}(L') = \dim \mathbb{P}(L'') = m - 1$ .  $\square$

By analogy with linear and affine transformations, there is a hope that we can describe a projective transformation unambiguously by how it maps a certain number of “sufficiently independent” points. As a first attempt, we may consider the points  $P_i = \langle e_i \rangle$  for  $i = 0, 1, \dots, n$ , where  $e_0, e_1, \dots, e_n$  is a basis of the space  $L$ . But this path does not lead to our goal, for there exist too many distinct transformations taking each point  $P_i$  into itself. Indeed, such are all the transformations of the form  $\mathbb{P}(\mathcal{A})$  if  $\mathcal{A}(e_i) = \lambda_i e_i$  with arbitrary  $\lambda_i \neq 0$ , that is, in other words, if  $\mathcal{A}$  has, in the basis  $e_0, e_1, \dots, e_n$ , the matrix

$$A = \begin{pmatrix} \lambda_0 & 0 & \dots & 0 \\ 0 & \lambda_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

In this case,  $\langle \mathcal{A}(e_i) \rangle = \langle e_i \rangle$  for all  $i = 0, 1, \dots, n$ . However, the image of an arbitrary vector

$$e = \alpha_0 e_0 + \alpha_1 e_1 + \dots + \alpha_n e_n$$

is equal to

$$\mathcal{A}(e) = \alpha_0 \lambda_0 \mathcal{A}(e_0) + \alpha_1 \lambda_1 \mathcal{A}(e_1) + \dots + \alpha_n \lambda_n \mathcal{A}(e_n),$$

and this vector is already not proportional to  $e$  unless all  $\lambda_i$  are identical. Thus even knowing how the transformation  $\mathbb{P}(\mathcal{A})$  maps the points  $P_0, P_1, \dots, P_n$ , we are not yet able to determine it uniquely. But it turns out that the addition of one more point (under some weak assumptions) describes the transformation uniquely. For this, we need to introduce a new concept.

**Definition 9.13** In the  $n$ -dimensional projective space  $\mathbb{P}(L)$ ,  $n + 2$  points

$$P_0, P_1, \dots, P_n, P_{n+1} \quad (9.17)$$

are said to be *independent* if no  $n + 1$  of them lie in a subspace of dimension less than  $n$ .

For example, four points in the projective plane are independent if no three of them are collinear.

Let us explore what the condition of independence means if to the point  $P_i$  there corresponds the line  $\langle e_i \rangle$ ,  $i = 0, \dots, n + 1$ . Since by definition, the points  $P_0, P_1, \dots, P_n$  do not lie in a subspace of dimension less than  $n$ , it follows that the vectors  $e_0, e_1, \dots, e_n$  do not lie in a subspace of dimension less than  $n + 1$ , that is, they are linearly independent, and this means that they constitute a basis of the space  $L$ . Thus the vector  $e_{n+1}$  is a linear combination of these vectors:

$$e_{n+1} = \alpha_0 e_0 + \alpha_1 e_1 + \dots + \alpha_n e_n. \quad (9.18)$$

If some scalar  $\alpha_i$  is equal to 0, then from (9.18), it follows that the vector  $e_{n+1}$  lies in the subspace  $L' = \langle e_0, \dots, \check{e}_i, \dots, e_n \rangle$ , where the sign  $\check{\phantom{x}}$  indicates the omission of the corresponding vector. Consequently, the vectors  $e_0, \dots, \check{e}_i, \dots, e_n, e_{n+1}$  lie in a subspace  $L'$  whose dimension does not exceed  $n$ . But this means that the points  $P_0, \dots, \check{P}_i, \dots, P_n, P_{n+1}$  lie in the projective space  $\mathbb{P}(L')$ , and moreover,  $\dim \mathbb{P}(L') \leq n - 1$ , that is, they are dependent.

Let us show that for the independence of points (9.17), it suffices that in the decomposition (9.18), all coefficients  $\alpha_i$  be nonzero. Let the vectors  $e_0, e_1, \dots, e_n$  form a basis of the space  $L$ , while the vector  $e_{n+1}$  is a linear combination (9.18) of them such that all the  $\alpha_i$  are nonzero. Let us show that then, the points (9.17) are independent. If this were not the case, then some  $n + 1$  vectors from among  $e_0, e_1, \dots, e_{n+1}$  of the space  $L$  would lie in a subspace of dimension not greater than  $n$ . This cannot be the vectors  $e_0, e_1, \dots, e_n$ , since by assumption, they constitute a basis of  $L$ . So let it be the vectors  $e_0, \dots, \check{e}_i, \dots, e_n, e_{n+1}$  for some  $i < n + 1$ , and their linear dependence is expressed by the equality

$$\lambda_0 e_0 + \dots + \lambda_{i-1} e_{i-1} + \lambda_{i+1} e_{i+1} + \dots + \lambda_{n+1} e_{n+1} = \mathbf{0},$$

where  $\lambda_{n+1} \neq 0$ , since the vectors  $e_0, e_1, \dots, e_n$  are linearly independent. From this, it follows that the vector  $e_{n+1}$  is a linear combination of the vectors  $e_0, \dots, \check{e}_i, \dots, e_n$ . But this contradicts the condition that in the expression (9.18), all the  $\alpha_i$  are nonzero, since the vectors  $e_0, e_1, \dots, e_n$  form a basis of the space  $L$ , and the decomposition (9.18) for an arbitrary vector  $e_{n+1}$  uniquely determines its coordinates  $\alpha_i$ .

Thus,  $n + 2$  independent points (9.17) are always obtained from  $n + 1$  points  $P_i = \langle e_i \rangle$  whose corresponding vectors  $e_i$  form a basis of the space  $L$  by the addition of one more point  $P = \langle e \rangle$  for which the vector  $e$  is a linear combination of the vectors  $e_i$  with all nonzero coefficients.

We can now formulate our main result.

**Theorem 9.14** *Let*

$$P_0, P_1, \dots, P_n, P_{n+1}; \quad P'_0, P'_1, \dots, P'_n, P'_{n+1} \quad (9.19)$$

*be two systems of independent points of the projective space  $\mathbb{P}(L)$  of dimension  $n$ . Then there exists a projective transformation taking the point  $P_i$  to  $P'_i$  for all  $i = 0, 1, \dots, n+1$ , and moreover, it is unique.*

*Proof* We shall use the interpretation of the property of independence of points obtained above. Let points  $P_i$  correspond to the lines  $\langle e_i \rangle$ , and let the points  $P'_i$  correspond to the lines  $\langle e'_i \rangle$ . We may assume that the vectors  $e_0, \dots, e_n$  and the vectors  $e'_0, \dots, e'_n$  are bases of an  $(n+1)$ -dimensional subspace of  $L$ . Then as we know, for every collection of nonzero scalars  $\lambda_0, \dots, \lambda_n$ , there exists (and it is unique) a nonsingular linear transformation  $\mathcal{A} : L \rightarrow L$  mapping  $e_i$  to  $\lambda_i e'_i$  for all  $i = 0, 1, \dots, n$ .

By definition, for such a transformation  $\mathcal{A}$ , we have  $\mathbb{P}(\mathcal{A})(P_i) = P'_i$  for all  $i = 0, 1, \dots, n$ . Since  $\dim L = n+1$ , we have the relationships

$$e_{n+1} = \alpha_0 e_0 + \alpha_1 e_1 + \dots + \alpha_n e_n, \quad e'_{n+1} = \alpha'_0 e'_0 + \alpha'_1 e'_1 + \dots + \alpha'_n e'_n. \quad (9.20)$$

From the condition of independence of both collections of points (9.19), it follows that in the representations (9.20), all the coefficients  $\alpha_i$  and  $\alpha'_i$  are nonzero. Applying the transformation  $\mathcal{A}$  to both sides of the first relationship in (9.20), taking into account the equalities  $\mathcal{A}(e_i) = \lambda_i e'_i$ , we obtain

$$\mathcal{A}(e_{n+1}) = \alpha_0 \lambda_0 e'_0 + \alpha_1 \lambda_1 e'_1 + \dots + \alpha_n \lambda_n e'_n. \quad (9.21)$$

After setting the scalars  $\lambda_i$  equal to  $\alpha'_i \alpha_i^{-1}$  for all  $i = 0, 1, \dots, n$  and substituting them into the relationship (9.21), taking into account the second equality of formula (9.20), we obtain that  $\mathcal{A}(e_{n+1}) = e'_{n+1}$ , that is,  $\mathbb{P}(\mathcal{A})(P_{n+1}) = P'_{n+1}$ .

The uniqueness of the projective transformation  $\mathbb{P}(\mathcal{A})$  that we have obtained follows from its construction.  $\square$

For example, for  $n = 1$ , the space  $\mathbb{P}(L)$  is the projective line. Three points  $P_0, P_1, P_2$  are independent if and only if they are distinct. We see that any three distinct points on the projective line can be mapped into three other distinct points by a unique projective transformation.

Let us now consider how a projective transformation can be given in coordinate form. In homogeneous coordinates (9.2), the stipulation of a projective transformation  $\mathbb{P}(\mathcal{A})$  in fact coincides with that of a nonsingular linear transformation  $\mathcal{A}$ , and indeed, the homogeneous coordinates of a point  $A \in \mathbb{P}(L)$  coincide with the coordinates of the vector  $x$  from (9.1) that determines the line  $\langle x \rangle$  corresponding to the point  $A$ . Using formula (3.25), we obtain for the homogeneous coordinates  $\beta_i$  of the point  $\mathbb{P}(\mathcal{A})(A)$  the following expressions in homogeneous coordinates  $\alpha_i$  of the



Thus the vector  $\mathbf{y} - \mathbf{x}$  has, in the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , coordinates

$$x_1 = \frac{\alpha_1}{\alpha_0}, \quad \dots, \quad x_n = \frac{\alpha_n}{\alpha_0}.$$

We shall now consider a nonsingular linear transformation  $\mathcal{A} : \mathbb{L} \rightarrow \mathbb{L}$  and the associated projective transformation  $\mathbb{P}(\mathcal{A})$ , given by formulas (9.22). It takes a point  $A$  with homogeneous coordinates  $\alpha_i$  to a point  $B$  with homogeneous coordinates  $\beta_i$ . In order to obtain in both cases inhomogeneous coordinates in the subset  $V_0$ , it is necessary, by formula (9.6), to divide all the coordinates by the coordinate with index 0. Thus we obtain that a point with inhomogeneous coordinates  $x_i = \frac{\alpha_i}{\alpha_0}$  is mapped to the point with inhomogeneous coordinates  $y_i = \frac{\beta_i}{\beta_0}$ , that is, taking into account (9.22), we obtain the expressions

$$y_i = \frac{a_{i0} + a_{i1}x_1 + \dots + a_{in}x_n}{a_{00} + a_{01}x_1 + \dots + a_{0n}x_n}, \quad i = 1, \dots, n. \quad (9.24)$$

In other words, in inhomogeneous coordinates, a projective transformation can be written in terms of the linear fractional formulas (9.24) with a common denominator for all  $y_i$ . It is not defined at points where this denominator becomes zero, and these are the “points at infinity,” that is, points of the projective hyperplane  $\mathbb{P}(\mathbb{L}_0)$  with equation  $\beta_0 = 0$ .

Let us consider projective transformations mapping “points at infinity” to “points at infinity” and consequently, “finite points” to “finite points.” This means that the equality  $\beta_0 = 0$  is possible only for  $\alpha_0 = 0$ , that is, taking into account formula (9.22), the equality

$$a_{00}\alpha_0 + a_{01}\alpha_1 + a_{02}\alpha_2 + \dots + a_{0n}\alpha_n = 0$$

is possible only for  $\alpha_0 = 0$ . Obviously, this latter condition is equivalent to the conditions  $a_{0i} = 0$  for all  $i = 1, \dots, n$ . In this case, the common denominator of the linear fractional formulas (9.24) reduces to the constant  $a_{00}$ . From the nonsingularity of the transformation  $\mathcal{A}$ , it follows that  $a_{00} \neq 0$ , and we can divide the numerators in equalities (9.24) by  $a_{00}$ . We then obtain precisely the formulas for affine transformations (8.17). Thus affine transformations are special cases of projective transformations, namely, those that take the set of “points at infinity” to itself.

*Example 9.15* In the case  $\dim \mathbb{P}(\mathbb{L}) = 1$ , the projective line  $\mathbb{P}(\mathbb{L})$  has a single inhomogeneous coordinate, and formula (9.24) assumes the form

$$y = \frac{a + bx}{c + dx}, \quad ad - bc \neq 0.$$

Transformations of the “finite part” of the projective line ( $x \neq \infty$ ) are affine and have the form  $y = \alpha + \beta x$ , where  $\beta \neq 0$ .

### 9.3 The Cross Ratio

Let us recall that in Sect. 8.2, we defined the affine ratio  $(A, B, C)$  among three collinear points of an affine space, and then, in Sect. 8.3, it was proved (Theorem 8.28) that the affine ratio  $(A, B, C)$  among three collinear points does not change under a nonsingular affine transformation. In projective spaces, the notion of a relationship among three collinear points cannot be given a natural analogue. This is the result of the following assertion.

**Theorem 9.16** *Let  $A_1, B_1, C_1$  and  $A_2, B_2, C_2$  be two triples of points in a projective space satisfying the following conditions:*

- (a) *The three points in each triple are distinct.*
- (b) *The points in each triple are collinear (one line for each triple).*

*Then there exists a projective transformation taking one triple into the other.*

*Proof* Let us denote the line on which the three points  $A_i, B_i, C_i$  lie by  $l_i$ , where  $i = 1, 2$ . Points  $A_1, B_1, C_1$  are independent on  $l_1$ , and the points  $A_2, B_2, C_2$  are independent on  $l_2$ . Let the point  $A_i$  be determined by the line  $\langle e_i \rangle$ , point  $B_i$  by the line  $\langle f_i \rangle$ , point  $C_i$  by the line  $\langle g_i \rangle$ , and line  $l_i$  by the two-dimensional space  $L_i$ ,  $i = 1, 2$ . They are all contained in the space  $L$  that determines our projective space. Repeating the proof of Theorem 9.14 verbatim, we shall construct an isomorphism  $\mathcal{A}' : L_1 \rightarrow L_2$  taking the lines  $\langle e_1 \rangle, \langle f_1 \rangle, \langle g_1 \rangle$  to the lines  $\langle e_2 \rangle, \langle f_2 \rangle, \langle g_2 \rangle$  respectively. Let us represent the space  $L$  in the form of two decompositions:

$$L = L_1 \oplus L'_1, \quad L = L_2 \oplus L'_2.$$

It is obvious that  $\dim L'_1 = \dim L'_2 = \dim L - 2$ , and therefore, the spaces  $L'_1$  and  $L'_2$  are isomorphic. We shall choose some isomorphism  $\mathcal{A}'' : L'_1 \rightarrow L'_2$  and define a transformation  $\mathcal{A} : L \rightarrow L$  as  $\mathcal{A}'$  on  $L_1$  and as  $\mathcal{A}''$  on  $L'_1$ , while for arbitrary vectors  $x \in L$ , we shall use the decomposition  $x = x_1 + x'_1$ ,  $x_1 \in L_1$ ,  $x'_1 \in L'_1$ , to define  $\mathcal{A}(x) = \mathcal{A}'(x_1) + \mathcal{A}''(x'_1)$ . It is easy to see that  $\mathcal{A}$  is a nonsingular linear transformation, and the projective transformation  $\mathbb{P}(\mathcal{A})$  takes the triple of points  $A_1, B_1, C_1$  to  $A_2, B_2, C_2$ .  $\square$

Analogously to the fact that for a triple of collinear points  $A, B, C$  of an affine space, there is an associated number  $(A, B, C)$  that is unchanged under every nonsingular affine transformation, in a projective space we can associate with a *quadruple* of collinear points  $A_1, A_2, A_3, A_4$  a number that does not change under projective transformations. This number is denoted by  $(A_1, A_2, A_3, A_4)$  and is called the *cross* or *anharmonic ratio* of these four points. We now turn to its definition.

Let us consider first the projective line  $l = \mathbb{P}(L)$ , where  $\dim L = 2$ . Four arbitrary points  $A_1, A_2, A_3, A_4$  on  $l$  correspond to four lines  $\langle a_1 \rangle, \langle a_2 \rangle, \langle a_3 \rangle, \langle a_4 \rangle$  lying in the plane  $L$ . In the plane  $L$ , let us choose a basis  $e_1, e_2$  and consider the decomposition

of the vectors  $\mathbf{a}_i$  in this basis:  $\mathbf{a}_i = x_i \mathbf{e}_1 + y_i \mathbf{e}_2$ ,  $i = 1, \dots, 4$ . The coordinates of the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_4$  can be written as the columns of the matrix

$$M = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \end{pmatrix}.$$

Consider the following question: how do the minors of order 2 of the matrix  $M$  change under a transition to another basis  $\mathbf{e}'_1, \mathbf{e}'_2$  of the plane  $L$ ? Let us denote by  $[\alpha_i]$  and  $[\alpha'_i]$  the columns of the coordinates of the vector  $\mathbf{a}_i$  in the bases  $(\mathbf{e}_1, \mathbf{e}_2)$  and  $(\mathbf{e}'_1, \mathbf{e}'_2)$  respectively:

$$[\alpha_i] = \begin{pmatrix} x_i \\ y_i \end{pmatrix}, \quad [\alpha'_i] = \begin{pmatrix} x'_i \\ y'_i \end{pmatrix}.$$

By formula (3.36) for changing coordinates, they are related by  $[\alpha] = C[\alpha']$ , where  $C$  is the transition matrix from the basis  $\mathbf{e}'_1, \mathbf{e}'_2$  to the basis  $\mathbf{e}_1, \mathbf{e}_2$ . From this it follows that

$$\begin{pmatrix} x_i & x_j \\ y_i & y_j \end{pmatrix} = C \cdot \begin{pmatrix} x'_i & x'_j \\ y'_i & y'_j \end{pmatrix}$$

for any choice of indices  $i$  and  $j$ , and by the theorem on multiplication of determinants, we obtain

$$\begin{vmatrix} x_i & x_j \\ y_i & y_j \end{vmatrix} = |C| \cdot \begin{vmatrix} x'_i & x'_j \\ y'_i & y'_j \end{vmatrix},$$

where  $|C| \neq 0$ . This means that for any three indices  $i, j, k$ , the relation

$$\frac{\begin{vmatrix} x_i & x_j \\ y_i & y_j \end{vmatrix}}{\begin{vmatrix} x_i & x_k \\ y_i & y_k \end{vmatrix}} = \frac{\begin{vmatrix} x'_i & x'_j \\ y'_i & y'_j \end{vmatrix}}{\begin{vmatrix} x'_i & x'_k \\ y'_i & y'_k \end{vmatrix}} \quad (9.25)$$

is unaltered under a change of basis (we assume now that both determinants, in the numerator and denominator, are nonzero). Thus relationship (9.25) determines a number  $(\mathbf{a}_i, \mathbf{a}_j, \mathbf{a}_k)$  depending on the three vectors  $\mathbf{a}_i, \mathbf{a}_j, \mathbf{a}_k$  but not on the choice of basis in  $L$ .

However, this is not yet what we promised: the points  $A_i$  indeed determine the lines  $\langle \mathbf{a}_i \rangle$ , but not the vectors  $\mathbf{a}_i$ . We know that the vector  $\mathbf{a}'_i$  determines the same line as the vector  $\mathbf{a}_i$  if and only if  $\mathbf{a}'_i = \lambda_i \mathbf{a}_i$ ,  $\lambda_i \neq 0$ . Therefore, if in expression (9.25) we replace the coordinates of the vectors  $\mathbf{a}_i, \mathbf{a}_j, \mathbf{a}_k$  with the coordinates of the proportional vectors  $\mathbf{a}'_i, \mathbf{a}'_j, \mathbf{a}'_k$ , then its numerator will be multiplied by  $\lambda_i \lambda_j$ , while its denominator will be multiplied by  $\lambda_i \lambda_k$ , with the result that the entire expression (9.25) will be multiplied by the number  $\lambda_j \lambda_k^{-1}$ , which means that it will change.

However, if we now consider the expression

$$DV(A_1, A_2, A_3, A_4) = \frac{\begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix} \cdot \begin{vmatrix} x_2 & x_4 \\ y_2 & y_4 \end{vmatrix}}{\begin{vmatrix} x_1 & x_4 \\ y_1 & y_4 \end{vmatrix} \cdot \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix}}, \quad (9.26)$$

then as our previous reasoning demonstrates, it will depend neither on the choice of basis of the plane  $L$  nor on the choice of vectors  $\mathbf{a}_i$  on the lines  $\langle \mathbf{a}_i \rangle$ , but will be determined only by the four points  $A_1, A_2, A_3, A_4$  on the projective line  $l$ . It is expression (9.26) that is called the *cross ratio* of these four points.

Let us write the expression for  $DV(A_1, A_2, A_3, A_4)$  assuming that homogeneous coordinates have been introduced on the projective line  $l$ . Let us begin with the formula written in the homogeneous coordinates  $(x : y)$ . We shall now consider the points  $A_i$  “finite” points of  $l$ , that is, we assume that  $y_i \neq 0$  for all  $i = 1, \dots, 4$ , and we set  $t_i = x_i / y_i$ ; these will be the coordinates of the point  $A_i$  in the “affine part” of the projective line  $l$ . Then we obtain

$$\begin{vmatrix} x_i & x_j \\ y_i & y_j \end{vmatrix} = y_i y_j \cdot \begin{vmatrix} t_i & t_j \\ 1 & 1 \end{vmatrix} = y_i y_j (t_i - t_j).$$

Substituting these expressions into formula (9.26), we see that all the  $y_i$  cancel, and as a result, we obtain the expression

$$DV(A_1, A_2, A_3, A_4) = \frac{(t_1 - t_3)(t_2 - t_4)}{(t_1 - t_4)(t_2 - t_3)}. \quad (9.27)$$

If we assume that all four points  $A_1, A_2, A_3, A_4$  lie in the “finite part” of the plane, then this means in particular that they belong to the affine part of the projective line  $l$  and have finite coordinates  $t_1, t_2, t_3, t_4$  on the projective line  $l$ . Taking into account formula (8.8) for the affine ratio of three points, we observe that then the expression for the cross ratio takes the form

$$DV(A_1, A_2, A_3, A_4) = \frac{(A_3, A_2, A_1)}{(A_4, A_2, A_1)}. \quad (9.28)$$

Equality (9.28) shows the connection between the cross ratio and the affine ratio introduced in Sect. 8.2.

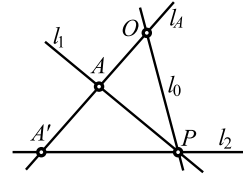
We have determined the cross ratio for four distinct points. In the case in which two of these points coincide, it is possible to define this ratio under some natural conventions (as we did for the affine ratio), setting the cross ratio in some cases equal to  $\infty$ . However, the cross ratio remains undefined if three of the four points coincide.

The above reasoning almost contains the proof of the following fundamental property of the cross ratio.

**Theorem 9.17** *The cross ratio of four collinear points in a projective space does not change under a projective transformation of the space.*



**Fig. 9.3** Perspective mapping



*Proof* Let  $A_1, A_2, A_3, A_4$  be four points lying on the line  $l'$  in some projective space  $\mathbb{P}(L)$ . They correspond to the four lines  $\langle a_1 \rangle, \langle a_2 \rangle, \langle a_3 \rangle, \langle a_4 \rangle$  of the space  $L$ , and the line  $l'$  corresponds to the two-dimensional subspace  $L' \subset L$ . Let  $\mathcal{A}$  be a nonsingular transformation of the space  $L$ , and  $\varphi = \mathbb{P}(\mathcal{A})$  the corresponding projective transformation of the space  $\mathbb{P}(L)$ . Then by Theorem 9.12,  $\varphi(l') = l''$  is another line in the projective space  $\mathbb{P}(L)$ ; it corresponds to the subspace  $\mathcal{A}(L') \subset L$  and contains the four points  $\varphi(A_1), \varphi(A_2), \varphi(A_3), \varphi(A_4)$ . Let the vectors  $e_1, e_2$  form a basis of  $L'$  and write the vectors  $a_i$  as  $a_i = x_i e_1 + y_i e_2$ ,  $i = 1, \dots, 4$ . Then the cross ratio  $DV(A_1, A_2, A_3, A_4)$  is defined by the formula (9.26).

On the other hand,  $\mathcal{A}(a_i) = x_i \mathcal{A}(e_1) + y_i \mathcal{A}(e_2)$ , and if we use the bases  $f_1 = \mathcal{A}(e_1)$  and  $f_2 = \mathcal{A}(e_2)$  of the subspace  $\mathcal{A}(L')$ , then the cross ratio

$$DV(\varphi(A_1), \varphi(A_2), \varphi(A_3), \varphi(A_4))$$

is defined by the same formula (9.26), since the coordinates of the vectors  $\mathcal{A}(a_i)$  in the basis  $f_1, f_2$  coincide with the coordinates of the vectors  $a_i$  in the basis  $e_1, e_2$ . But as we have already verified, the cross ratio depends neither on the choice of basis nor on the choice of vectors  $a_i$  that determine the lines  $\langle a_i \rangle$ . Therefore, it follows that

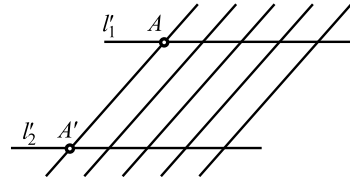
$$DV(A_1, A_2, A_3, A_4) = DV(\varphi(A_1), \varphi(A_2), \varphi(A_3), \varphi(A_4)). \quad \square$$

*Example 9.18* In a projective space  $\Pi$ , let us consider two lines  $l_1$  and  $l_2$  and a point  $O$  lying on neither of the lines. Let us connect an arbitrary point  $A \in l_1$  to the point  $O$  of the line  $l_A$ ; see Fig. 9.3. We shall denote the point of intersection of the lines  $l_A$  and  $l_2$  by  $A'$ . The mapping of the line  $l_1$  into  $l_2$  that to each point  $A \in l_1$  assigns the point  $A' \in l_2$  is called a *perspective mapping*.

Let us prove that there exists a projective transformation of the plane  $\Pi$  defining a perspective correspondence between the lines  $l_1$  and  $l_2$ . To this end, let us denote by  $l_0$  the line joining the point  $O$  and the point  $P = l_1 \cap l_2$ , and let us consider the set  $V = \Pi \setminus l_0$ . In other words, we shall consider  $l_0$  a “line at infinity” and the points of  $V$  will be considered “finite points” of the projective plane. Then on  $V$ , the perspective correspondence will be given by a bundle of parallel lines, since these lines in the “finite part” do not intersect; see Fig. 9.4.

More precisely, this bundle defines a mapping of the “finite parts”  $l'_1$  and  $l'_2$  of the lines  $l_1$  and  $l_2$ . From this it follows that in the affine plane  $V$ , the lines  $l'_1$  and  $l'_2$  are parallel, and the perspective correspondence between them is defined by an

**Fig. 9.4** A bundle of *parallel* lines



arbitrary translation  $\mathcal{T}_a$  by the vector  $a = \overrightarrow{AA'}$ , where  $A$  is an arbitrary point on the line  $l'_1$ , and  $A'$  is the point on the line  $l'_2$  corresponding to it under the perspective correspondence. As we saw above, every nonsingular affine transformation of an affine plane  $V$  is a projective mapping for  $\Pi$ , and this is even more obviously the case for a translation. This means that a perspective correspondence is defined by some projective transformation of the plane  $\Pi$ . Therefore, from Theorem 9.17, we deduce the following result.

**Theorem 9.19** *The cross ratio of four collinear points is preserved under a perspective correspondence.*

## 9.4 Topological Properties of Projective Spaces\*

The previous discussion in this chapter was related to a projective space  $\mathbb{P}(L)$ , where  $L$  was a finite-dimensional vector space over an arbitrary field  $\mathbb{K}$ . If our interest is in a particular field (for example,  $\mathbb{R}$  or  $\mathbb{C}$ ), then all the assertions we have proved remain valid, since we used only general algebraic notions (which derive from the definition of a field), and nowhere did we use, for example, properties of inequality or absolute value. Now let us say a few words about properties related to the notion of *convergence*, or as they are called, *topological* properties, of projective spaces. It makes sense to talk about them if, for example,  $L$  is a real or complex vector space, that is, the field in question is  $\mathbb{K} = \mathbb{R}$  or  $\mathbb{C}$ .

Let us begin by formulating the notion of *convergence* of a sequence of vectors  $x_1, x_2, \dots, x_k, \dots$  in a space  $L$  to a vector  $x$  of the same space. Let us choose in  $L$  an arbitrary basis  $e_0, e_1, \dots, e_n$  and let us write the vectors  $x_k$  and  $x$  in this basis:

$$x_k = \alpha_{k0}e_0 + \alpha_{k1}e_1 + \dots + \alpha_{kn}e_n, \quad x = \beta_0e_0 + \beta_1e_1 + \dots + \beta_ne_n.$$

We shall say that the sequence of vectors  $x_1, x_2, \dots, x_k, \dots$  *converges* to the vector  $x$  if the sequence of numbers

$$\alpha_{1i}, \alpha_{2i}, \dots, \alpha_{ki}, \dots \quad (9.29)$$

for fixed  $i$  converges to the number  $\beta_i$  as  $k \rightarrow \infty$  for each index  $i = 0, 1, \dots, n$  (in speaking about complex vector spaces, we assume that the reader is familiar with the notion of convergence of a sequence of complex numbers). The vector  $x$  is called, in this case, the *limit* of the sequence. From the formulas for changing coordinates

given in Sect. 3.4, it is easy to derive that the property of convergence does not depend on the basis in  $L$ . We shall write this convergence as  $\mathbf{x}_k \rightarrow \mathbf{x}$  as  $k \rightarrow \infty$ .

Let us move now from vectors to points of a projective space. In both cases that we are considering ( $\mathbb{K} = \mathbb{R}$  or  $\mathbb{C}$ ), there is a useful method of *normalizing* the homogeneous coordinates  $(x_0 : x_1 : \dots : x_n)$  defined, generally speaking, only up to multiplication by a common factor  $\lambda \neq 0$ . Since by definition, the equality  $x_i = 0$  for all  $i = 0, 1, \dots, n$  is impossible, we may choose a coordinate  $x_r$  for which  $|x_r|$  (the absolute value in  $\mathbb{R}$  or  $\mathbb{C}$ , respectively) assumes the greatest value, and setting  $\lambda = |x_r|$ , make the substitution  $y_i = \lambda^{-1}x_i$  for all  $i = 0, 1, \dots, n$ . Then, obviously,

$$(x_0 : x_1 : \dots : x_n) = (y_0 : y_1 : \dots : y_n),$$

and moreover,  $|y_r| = 1$  and  $|y_i| \leq 1$  for all  $i = 0, 1, \dots, n$ .

**Definition 9.20** A sequence of points  $P_1, P_2, \dots, P_k, \dots$  converges to the point  $P$  if on every line  $\langle \mathbf{e}_k \rangle$  that determines the point  $P_k$ , and on the line  $\langle \mathbf{e} \rangle$  determining the point  $P$ , it is possible to find nonnull vectors  $\mathbf{x}_k$  and  $\mathbf{x}$  such that  $\mathbf{x}_k \rightarrow \mathbf{x}$  as  $k \rightarrow \infty$ . This is written as  $P_k \rightarrow P$  as  $k \rightarrow \infty$ . The point  $P$  is called the *limit* of the sequence  $P_1, P_2, \dots, P_k, \dots$ .

We note that by assumption,  $\langle \mathbf{e}_k \rangle = \langle \mathbf{x}_k \rangle$  and  $\langle \mathbf{e} \rangle = \langle \mathbf{x} \rangle$ .

**Theorem 9.21** It is possible to choose from an arbitrary infinite sequence of points of a projective space a subsequence that converges to a point of the space.

*Proof* As we have seen, every point  $P$  of a projective space can be represented in the form  $P = \langle \mathbf{y} \rangle$ , where the vector  $\mathbf{y}$  has coordinates  $(y_0, y_1, \dots, y_n)$ , and moreover,  $\max |y_i| = 1$ .

It is proved in a course in real analysis that every bounded sequence of real numbers satisfies the assertion of Theorem 9.21. It is also very easy to prove the statement for a sequence of complex numbers. To obtain from this the assertion of the theorem, let us consider an infinite sequence of points  $P_1, P_2, \dots, P_k, \dots$  of the projective space  $\mathbb{P}(L)$ . Let us focus attention first on the sequence of zeroth (that is, having index 0) coordinates of the vectors  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \dots$  corresponding to these points. Suppose they are the numbers

$$\alpha_{10}, \alpha_{20}, \dots, \alpha_{k0}, \dots \quad (9.30)$$

As we noted above, we may assume that all  $|\alpha_{k0}|$  are less than or equal to 1. By the assertion from real analysis formulated above, from the sequence (9.30), we may choose a subsequence

$$\alpha_{n_1 0}, \alpha_{n_2 0}, \dots, \alpha_{n_k 0}, \dots, \quad (9.31)$$

converging to some number  $\beta_0$  that therefore also does not exceed 1 in absolute value. Let us now consider a subsequence of points  $P_{n_1}, P_{n_2}, \dots, P_{n_k}, \dots$  and of vectors  $\mathbf{x}_{n_1}, \mathbf{x}_{n_2}, \dots, \mathbf{x}_{n_k}, \dots$  with the same indices as those in the subsequence

(9.31). Let us focus attention on the first coordinate of these vectors. For them, clearly, it is also the case that  $|\alpha_{n_k 1}| \leq 1$ . This means that from the sequence

$$\alpha_{n_1 1}, \alpha_{n_2 1}, \dots, \alpha_{n_k 1}, \dots$$

we may choose a subsequence converging to some number  $\beta_1$ , and moreover, clearly  $|\beta_1| \leq 1$ .

Repeating this argument  $n + 1$  times, we obtain as a result, from the original sequence of vectors  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \dots$ , a subsequence  $\mathbf{x}_{m_1}, \mathbf{x}_{m_2}, \dots, \mathbf{x}_{m_k}, \dots$  converging to some vector  $\bar{\mathbf{x}} \in L$ , which, like every vector of this space, can be decomposed in terms of the basis  $\mathbf{e}_0, \mathbf{e}_1, \dots, \mathbf{e}_n$ , that is,

$$\bar{\mathbf{x}} = \beta_0 \mathbf{e}_0 + \beta_1 \mathbf{e}_1 + \dots + \beta_n \mathbf{e}_n.$$

This gives us the assertion of Theorem 9.21 if we ascertain that not all coordinates  $\beta_0, \beta_1, \dots, \beta_n$  of the vector  $\bar{\mathbf{x}}$  are equal to zero. But this follows from the fact that by construction, for each vector  $\mathbf{x}_{m_k}$  of the subsequence  $\mathbf{x}_{m_1}, \mathbf{x}_{m_2}, \dots, \mathbf{x}_{m_k}, \dots$ , a certain coordinate  $\alpha_{m_k i}$ ,  $i = 0, \dots, n$ , has absolute value equal to 1. Since there exists only a finite number of coordinates, and the number of vectors  $\mathbf{x}_{m_k}$  is infinite, there must be an index  $i$  such that among the coordinates  $\alpha_{m_k i}$ , infinitely many will have absolute value 1. On the other hand, by construction, the sequence  $\alpha_{m_1 i}, \alpha_{m_2 i}, \dots, \alpha_{m_k i}, \dots$  converges to the number  $\beta_i$ , which therefore must have absolute value equal to 1.  $\square$

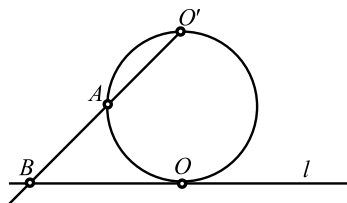
The property established in Theorem 9.21 is called *compactness*. It holds as well for every projective algebraic variety of a projective space (whether real or complex). We may formulate it as follows.

**Corollary 9.22** *In the case of a real or complex space, the points of a projective algebraic variety form a compact set.*

*Proof* Let the projective algebraic variety  $X$  be given by a system of equations (9.5), and let  $P_1, P_2, \dots, P_k, \dots$  be a sequence of points in  $X$ . By Theorem 9.21, there exists a subsequence of this sequence that converges to some point  $P$  of this space. It remains to prove that the point  $P$  belongs to the variety  $X$ . For this, it suffices to show that it can be represented in the form  $P = \langle \mathbf{u} \rangle$ , where the coordinates of the vector  $\mathbf{u}$  satisfy equations (9.5). But this follows at once from the fact that polynomials are continuous functions. Let  $F(x_0, x_1, \dots, x_n)$  be a polynomial (in this case, homogeneous; it is one of the polynomials  $F_i$  appearing in the system of equations (9.5)). We shall write it in the form  $F = F(\mathbf{x})$ , where  $\mathbf{x} \in L$ . Then from the convergence of the vectors  $\mathbf{x}_k \rightarrow \mathbf{x}$  as  $k \rightarrow \infty$  such that  $F(\mathbf{x}_k) = 0$  for all  $k$ , it follows that  $F(\mathbf{x}) = 0$ .  $\square$

For subsets of a finite-dimensional vector or affine space (whether real or complex), the property of compactness is related to their boundedness—more precisely,

**Fig. 9.5** The real projective line



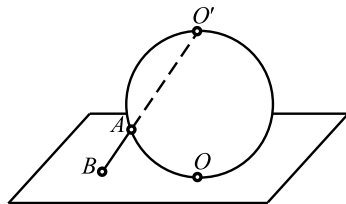
the property of boundedness follows from compactness. Thus while real and complex vector or affine spaces can be visualized as “extending unboundedly in all directions,” for projective spaces, such is not the case. But what does it mean to say “can be visualized”? In order to formulate this intuitive idea precisely, we shall introduce for the real and complex projective lines some simple geometric representations to which they are homeomorphic (see the relevant definition on p. xviii). This will allow us to give a precise meaning to the words that a given set “can be visualized.” Let us observe that the property of compactness established in Theorem 9.21 is unchanged under a transition from one set to another that is homeomorphic to it.

Let us begin with the simplest situation: a one-dimensional real projective space, that is, the *real projective line*. It consists of pairs  $(x_0 : x_1)$ , where  $x_0$  and  $x_1$  are considered only up to a common factor  $\lambda \neq 0$ . Those pairs for which  $x_0 \neq 0$  form an affine subset  $U$ , whose points are given by the single coordinate  $t = x_1/x_0$ , so that we may identify the set  $U$  with  $\mathbb{R}$ . Pairs for which  $x_0 = 0$  do not enter the set  $U$ , but they correspond to only one point  $(0 : 1)$  of the projective line, which we shall denote by  $(\infty)$ . Thus the real projective line can be represented in the form  $\mathbb{R} \cup (\infty)$ .

The convergence of points  $P_k \rightarrow Q$  as  $k \rightarrow \infty$  is defined in this case as follows. If points  $P_k \neq (\infty)$  correspond to the numbers  $t_k$ , and the point  $Q \neq (\infty)$  corresponds to the number  $t$ , then  $P_k = (\alpha_k : \beta_k)$  and  $Q = (\alpha : \beta)$ , where  $\beta_k/\alpha_k = t_k$ ,  $\alpha_k \neq 0$ , and  $\beta/\alpha = t$ ,  $\alpha \neq 0$ . The convergence  $P_k \rightarrow Q$  as  $k \rightarrow \infty$  in this case implies the convergence of the sequence of numbers  $t_k \rightarrow t$  as  $k \rightarrow \infty$ . In the case that  $P_k \rightarrow (\infty)$ , the convergence (in the previous notation) means that  $\alpha_k \rightarrow 0$ ,  $\beta_k \rightarrow 1$  as  $k \rightarrow \infty$ , from which it follows that  $t_k^{-1} \rightarrow 0$ , or equivalently,  $|t_k| \rightarrow \infty$  as  $k \rightarrow \infty$ .

We can graphically represent the real projective line by drawing a circle tangent to the horizontal line  $l$  at the point  $O$ ; see Fig. 9.5. Connecting the highest point  $O'$  of this circle with an arbitrary point  $A$  of the circle, we obtain a line that intersects  $l$  at some point  $B$ . We thereby obtain a bijection between points  $A \neq O'$  of the circle and all the points  $B$  of the line  $l$ . If we place the coordinate origin of the line  $l$  at the point  $O$  and associate with each point  $B \in l$  a number  $t \in \mathbb{R}$  resulting from a choice of some unit measure on the line  $l$  (that is, an arbitrary point of the line  $l$  different from  $O$  is given the value 1), then we obtain a bijection between numbers  $t \in \mathbb{R}$  and points  $A \neq O'$  of the circle. Then  $|t_k| \rightarrow \infty$  if and only if for the corresponding points  $A_k$  of the circle, we have the convergence  $A_k \rightarrow O'$ . Consequently, we obtain a bijection between points of the real projective line  $\mathbb{R} \cup (\infty)$  and *all* points of the circle that preserves the notion of convergence. Thus we have proved that the real

**Fig. 9.6** Stereographic projection of the *sphere* onto the *plane*



projective line is homeomorphic to the circle, which is usually denoted by  $S^1$  (the one-dimensional sphere).

An analogous argument can be applied to the complex projective line. It is represented in the form  $\mathbb{C} \cup (\infty)$ . On it, the convergence of a sequence of points  $P_k \rightarrow Q$  as  $k \rightarrow \infty$  in the case  $Q \neq (\infty)$  corresponds to convergence of a sequence of complex numbers  $z_k \rightarrow z$ , where  $z \in \mathbb{C}$ , while the convergence of the sequence of points  $P_k \rightarrow (\infty)$  corresponds to the convergence  $|z_k| \rightarrow \infty$  (here  $|z|$  denotes the modulus of the complex number  $z$ ).

For the graphical representation of the complex projective line, Riemann proposed the following method; see Fig. 9.6. The complex numbers are depicted in the usual way as points in a plane. Let us consider a sphere tangent to this plane at the origin  $O$ , which corresponds to the complex number  $z = 0$ . Through the highest point  $O'$  of the sphere and any other point  $A$  of the sphere there passes a line intersecting the complex plane at a point  $B$ , which represents some number  $z \in \mathbb{C}$ . This yields a bijection between numbers  $z \in \mathbb{C}$  and all the points of the sphere, with the exception of the point  $O'$ ; see Fig. 9.6. This correspondence is often called the *stereographic projection* of the sphere onto the plane. By associating the point  $(\infty)$  of the complex projective line with the point  $O'$  of the sphere, we obtain a bijection between the points of the complex projective line  $\mathbb{C} \cup (\infty)$  and *all* the points of the sphere. It is easy to see that convergence is preserved under this assignment. Thus the complex projective line is homeomorphic to the two-dimensional sphere in three-dimensional space, which is denoted by  $S^2$ .

In the sequel, we shall limit our consideration to projective spaces  $\mathbb{P}(L)$ , where  $L$  is a real vector space of some finite dimension, and we shall consider for such spaces the property of orientability. It is related to the concept of continuous deformation of a linear transformation, which was introduced in Sect. 4.4.

By definition, every projective transformation of a projective space  $\mathbb{P}(L)$  has the form  $\mathbb{P}(\mathcal{A})$ , where  $\mathcal{A}$  is a nonsingular linear transformation of the vector space  $L$ . Moreover, as we have seen, the linear transformation  $\mathcal{A}$  is determined by the projective transformation up to a replacement by  $\alpha\mathcal{A}$ , where  $\alpha$  is any nonzero number.

**Definition 9.23** A projective transformation is said to be *continuously deformable* into another if the first can be represented in the form  $\mathbb{P}(\mathcal{A}_1)$  and the second in the form  $\mathbb{P}(\mathcal{A}_2)$ , and the linear transformation  $\mathcal{A}_1$  is continuously deformable into  $\mathcal{A}_2$ .

Theorem 4.39 asserts that a linear transformation  $\mathcal{A}_1$  is continuously deformable into  $\mathcal{A}_2$  if and only if the determinants  $|\mathcal{A}_1|$  and  $|\mathcal{A}_2|$  have the same sign. What

happens under a replacement of  $\mathcal{A}$  by  $\alpha\mathcal{A}$ ? Let the projective space  $\mathbb{P}(L)$  have dimension  $n$ . Then the vector space  $L$  has dimension  $n+1$ , and  $|\alpha\mathcal{A}| = \alpha^{n+1}|\mathcal{A}|$ . If the number  $n+1$  is even, then it is always the case that  $\alpha^{n+1} > 0$ , and such a replacement does not change the sign of the determinant. In other words, in a projective space of odd dimension  $n$ , the sign of the determinant  $|\mathcal{A}|$  of a linear transformation  $\mathcal{A}$  is uniquely determined by the transformation  $\mathbb{P}(\mathcal{A})$ . This clearly yields the following result.

**Theorem 9.24** *In a projective space of odd dimension, a projective transformation  $\mathbb{P}(\mathcal{A}_1)$  is continuously deformable into  $\mathbb{P}(\mathcal{A}_2)$  if and only if the determinants  $|\mathcal{A}_1|$  and  $|\mathcal{A}_2|$  have the same sign.*

The same considerations can be applied to projective spaces of even dimension, but they lead to a different result.

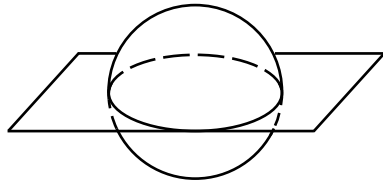
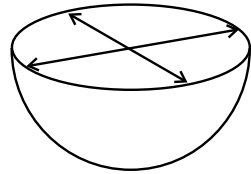
**Theorem 9.25** *In a projective space of even dimension, every projective transformation is continuously deformable into every other projective transformation.*

*Proof* Let us show that every projective transformation  $\mathbb{P}(\mathcal{A})$  is continuously deformable into the identity. If  $|\mathcal{A}| > 0$ , then this follows at once from Theorem 4.39. And if  $|\mathcal{A}| < 0$ , then the same theorem gives us that the transformation  $\mathcal{A}$  is continuously deformable into  $\mathcal{B}$ , which has matrix  $\begin{pmatrix} -1 & 0 \\ 0 & E_n \end{pmatrix}$ , where  $E_n$  is the identity matrix of order  $n$ . But  $\mathbb{P}(\mathcal{B}) = \mathbb{P}(-\mathcal{B})$ , and the transformation  $-\mathcal{B}$  has matrix  $\begin{pmatrix} 1 & 0 \\ 0 & -E_n \end{pmatrix}$ . Since in our case, the number  $n$  is even, it follows that  $|-E_n| = (-1)^n > 0$ , and by Theorem 4.38, the matrix  $\begin{pmatrix} 1 & 0 \\ 0 & -E_n \end{pmatrix}$  is continuously deformable into  $E_{n+1}$ , and consequently, the transformation  $-\mathcal{B}$  is continuously deformable into the identity. Thus the projective transformation  $\mathbb{P}(\mathcal{B})$  is continuously deformable into  $\mathbb{P}(\mathcal{E})$ , and this means by definition, that  $\mathbb{P}(\mathcal{A})$  is also continuously deformable into  $\mathbb{P}(\mathcal{E})$ .  $\square$

Expressing these facts in topological form, we may say that the set of projective transformations of the space  $\mathbb{P}^n$  of a given dimension has a single path-connected component if  $n$  is even, and two path-connected components if  $n$  is odd.

Theorems 9.24 and 9.25 show that the properties of projective spaces of even and odd dimension are radically different. We encounter this for the first time in the case of the projective plane. It differs from the vector (or Euclidean) plane in that it has not two, but only one orientation. It is the same with projective spaces of arbitrary even dimension. We saw in Sect. 4.4 that the orientation of the affine plane can be interpreted as a choice of direction of motion around a circle. Theorem 9.25 shows that in the projective plane, this is already not the case—the continuous motion in a given direction around a circle in the projective plane can be transformed into motion in the opposite direction. This is possible only because our deformation at a certain moment “passes through infinity,” which is impossible in the affine plane.

This property can be presented graphically using the following construction, which is applicable to real projective spaces of arbitrary dimension.

**Fig. 9.7** A model of the projective plane**Fig. 9.8** Identification of points

Let us assume that the vector space  $L$  defining our projective space  $\mathbb{P}(L)$  is a Euclidean space, and let us consider in this space the sphere  $S$ , defined by the equality  $|\mathbf{x}| = 1$ . Every line  $\langle \mathbf{x} \rangle$  of the space  $L$  intersects the sphere  $S$ . Indeed, such a line consists of vectors of the form  $\alpha \mathbf{x}$ , where  $\alpha \in \mathbb{R}$ , and the condition  $\alpha \mathbf{x} \in S$  means that  $|\alpha \mathbf{x}| = 1$ . Since  $|\alpha \mathbf{x}| = |\alpha| \cdot |\mathbf{x}|$  and  $\mathbf{x} \neq \mathbf{0}$ , we may set  $|\alpha| = |\mathbf{x}|^{-1}$ . With this choice, the number  $\alpha$  is determined up to sign, or in other words, there exist two vectors,  $\mathbf{e}$  and  $-\mathbf{e}$ , belonging to the line  $\langle \mathbf{x} \rangle$  and to the sphere  $S$ . Thus associating with each vector  $\mathbf{e} \in S$  the line  $\langle \mathbf{x} \rangle$  of the projective space, we obtain the mapping  $f : S \rightarrow \mathbb{P}(L)$ . The previous reasoning shows that the image of  $f$  is the entire space  $\mathbb{P}(L)$ . However, this mapping  $f$  is not a bijection, since two points of the sphere  $S$  pass through one point  $P \in \mathbb{P}(L)$ , corresponding to the line  $\langle \mathbf{x} \rangle$ , namely, the vectors  $\mathbf{e}$  and  $-\mathbf{e}$ . This property is expressed by saying that the projective space is obtained from the sphere  $S$  via the *identification* of its antipodal points.

Let us apply this to the case of the projective plane, that is, we shall suppose that  $\dim \mathbb{P}(L) = 2$ . Then  $\dim L = 3$ , and the sphere  $S$  contained in three-dimensional space is the sphere  $S^2$ . Let us decompose it into two equal parts by a horizontal plane; see Fig. 9.7.

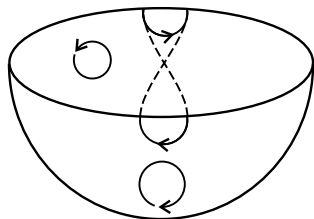
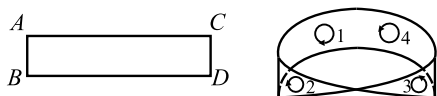
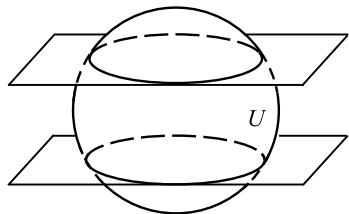
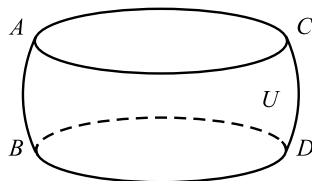
Each point of the upper hemisphere is diametrically opposite some point on the lower hemisphere, and we can map the upper hemisphere onto the projective plane  $\mathbb{P}(L)$  by representing each point  $P \in \mathbb{P}(L)$  in the form  $\langle \mathbf{e} \rangle$ , where  $\mathbf{e}$  is a vector of the upper hemisphere.

However, this correspondence will not be a bijection, since antipodal points on the boundary of the hemisphere will be joined together, that is, they correspond to a single point; see Fig. 9.8. This is expressed by saying that the projective plane is obtained by identifying antipodal points of the boundary of the hemisphere.

Let us now consider a moving circle with a given direction of rotation; see Fig. 9.9. In the figure is shown that when the moving circle intersects the boundary of the hemisphere, the direction of rotation changes to its opposite.

This property is expressed by saying that the projective plane is a *one-sided* surface (while the sphere in three-dimensional space and other familiar surfaces are *two-sided*). This property of the projective plane was studied by Möbius. He



**Fig. 9.9** Motion of a circle**Fig. 9.10** Möbius strip**Fig. 9.11** Partition of the sphere**Fig. 9.12** The central part of the sphere

presented an example of a one-sided surface that is now known as the *Möbius strip*. It can be constructed by cutting from a sheet of paper the rectangle  $ABDC$  (Fig. 9.10, left) and gluing together its opposite sides  $AB$  and  $CD$ , after rotating  $CD$  by  $180^\circ$ . The one-sided surface thus obtained is shown in the right-hand picture of Fig. 9.10, where is also shown the continuous deformation of the circle (stages  $1 \rightarrow 2 \rightarrow 3 \rightarrow 4$ ), changing the direction of rotation to it opposite.

The Möbius strip also has a direct relationship to the projective plane. Namely, let us visualize this plane as the sphere  $S^2$ , in which antipodal points are identified. Let us divide the sphere into three parts by intersecting it with two parallel planes that pass above and below the equator. As a result, the sphere is partitioned into a central part  $U$  and two “caps” above and below; see Fig. 9.11.

Let us begin by studying the central section  $U$ . For each point of  $U$ , its antipodal point is also contained in  $U$ . Let us divide  $U$  into two halves—front and back—by a vertical plane intersecting  $U$  in the arcs  $AB$  and  $CD$ ; see Fig. 9.12.

We may combine the front half ( $U'$ ) with the rectangle  $ABDC$  in Fig. 9.10. Every point of the central section  $U$  either itself belongs to the front half or else has an antipodal point that belongs to the front half, of which there is only one, except

for the points of the segments  $AB$  and  $CD$ . In order to obtain only one of the two antipodal points of these segments, we must glue these segments together exactly as is done in Fig. 9.10. Thus the Möbius strip is homeomorphic to the part  $U'$  of the projective plane. To obtain the remaining part  $V = \mathbb{P}(L) \setminus U'$ , we have to consider the “caps” on the sphere; see Fig. 9.11. For every point in a cap, its antipodal point lies in the other cap. This means that by identifying antipodal points, it suffices to consider only one cap, for example the upper one. This cap is homeomorphic to a disk: to see this, it suffices simply to project it onto the horizontal plane. Clearly, the boundary of the upper cap is identified with the boundary of the central part of the sphere. Thus the projective plane is homeomorphic to the surface obtained by gluing a circle to the Möbius strip in such a way that its boundary is identified with the boundary of the Möbius strip (it is easily verified that the boundary of the Möbius strip is a circle).

## Chapter 10

# The Exterior Product and Exterior Algebras

### 10.1 Plücker Coordinates of a Subspace

The fundamental idea of analytic geometry, which goes back to Fermat and Descartes, consists in the fact that every point of the two-dimensional plane or three-dimensional space is defined by its coordinates (two or three, respectively). Of course, there must also be present a particular choice of coordinate system. In this course, we have seen that this very principle is applicable to many spaces of more general types: vector spaces of arbitrary dimension, as well as Euclidean, affine, and projective spaces. In this chapter, we shall show that it can be applied to the study of vector subspaces  $M$  of fixed dimension  $m$  in a given vector space  $L$  of dimension  $n \geq m$ . Since there is a bijection between the  $m$ -dimensional subspaces  $M \subset L$  and  $(m - 1)$ -dimensional projective subspaces  $\mathbb{P}(M) \subset \mathbb{P}(L)$ , we shall therefore also obtain a description of the projective subspaces of fixed dimension of a projective space with the aid of “coordinates” (certain collections of numbers).

The case of *points* of a projective space (subspaces of dimension 0) was already analyzed in the previous chapter: they are given by homogeneous coordinates. The same holds in the case of *hyperplanes* of a projective space  $\mathbb{P}(L)$ : they correspond to the points of the dual space  $\mathbb{P}(L^*)$ . The simplest case in which the problem is not reduced to these two cases given above is the set of projective lines in three-dimensional projective space. Here a solution was proposed by Plücker. And therefore, in the most general case, the “coordinates” corresponding to the subspace are called *Plücker coordinates*. Following the course of history, we shall begin in Sects. 10.1 and 10.2 by describing these using some coordinate system, and then investigate the construction we have introduced in an invariant way, in order to determine which of its elements depend on the choice of coordinate system and which do not.

Therefore, we now assume that some basis has been chosen in the vector space  $L$ . Since  $\dim L = n$ , every vector  $\mathbf{a} \in L$  has in this basis  $n$  coordinates. Let us consider a subspace  $M \subset L$  of dimension  $m \leq n$ . Let us choose an arbitrary basis  $\mathbf{a}_1, \dots, \mathbf{a}_m$  of the subspace  $M$ . Then  $M = \langle \mathbf{a}_1, \dots, \mathbf{a}_m \rangle$ , and the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m$  are linearly

independent. The vector  $\mathbf{a}_i$  has, in the chosen basis of the space  $L$ , coordinates  $a_{i1}, \dots, a_{in}$  ( $i = 1, \dots, m$ ), which we can arrange in the form of a matrix  $M$  of type  $(m, n)$ , writing them in row form:

$$M = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}. \quad (10.1)$$

The condition that the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m$  are linearly independent means that the rank of the matrix  $M$  is equal to  $m$ , that is, one of its minors of order  $m$  is nonzero. Since the number of rows of the matrix  $M$  is equal to  $m$ , a minor of order  $m$  is uniquely defined by the indices of its columns. Let us denote by  $M_{i_1, \dots, i_m}$  the minor consisting of columns with indices  $i_1, \dots, i_m$ , which assume the various values from 1 to  $n$ .

We know that not all of the minors  $M_{i_1, \dots, i_m}$  can be equal to zero at the same time. Let us examine how they depend on the choice of basis  $\mathbf{a}_1, \dots, \mathbf{a}_m$  in  $M$ . If  $\mathbf{b}_1, \dots, \mathbf{b}_m$  is some other basis of this subspace, then

$$\mathbf{b}_i = b_{i1}\mathbf{a}_1 + \cdots + b_{im}\mathbf{a}_m, \quad i = 1, \dots, m.$$

Since the vectors  $\mathbf{b}_1, \dots, \mathbf{b}_m$  are linearly independent, the determinant  $|(b_{ij})|$  is nonzero. Let us set  $c = |(b_{ij})|$ . If  $M'_{i_1, \dots, i_m}$  is a minor of the matrix  $M'$ , constructed analogously to  $M$  using the vectors  $\mathbf{b}_1, \dots, \mathbf{b}_m$ , then by formula (3.35) and Theorem 2.54 on the determinant of a product of matrices, we have the relationship

$$M'_{i_1, \dots, i_m} = c M_{i_1, \dots, i_m}. \quad (10.2)$$

The numbers  $M_{i_1, \dots, i_m}$  that we have determined are not independent. Namely, if the unordered collection of numbers  $j_1, \dots, j_m$  coincides with  $i_1, \dots, i_m$  (that is, comprises the same numbers, perhaps arranged in a different order), then as we saw in Sect. 2.6, we have the relationship

$$M_{j_1, \dots, j_m} = \pm M_{i_1, \dots, i_m}, \quad (10.3)$$

where the sign  $+$  or  $-$  appears depending on whether the number of transpositions necessary to effect the passage from the collection  $(i_1, \dots, i_m)$  to  $(j_1, \dots, j_m)$  is even or odd. In other words, the function  $M_{i_1, \dots, i_m}$  of  $m$  arguments  $i_1, \dots, i_m$  assuming the values  $1, \dots, n$  is antisymmetric.

In particular, we may take as the collection  $(j_1, \dots, j_m)$  the arrangement of the numbers  $i_1, \dots, i_m$  such that  $i_1 < i_2 < \cdots < i_m$ , and the corresponding minor  $M_{j_1, \dots, j_m}$  will coincide with either  $M_{i_1, \dots, i_m}$  or  $-M_{i_1, \dots, i_m}$ . In view of this, in the original notation, we shall assume that  $i_1 < i_2 < \cdots < i_m$ , and we shall set

$$p_{i_1, \dots, i_m} = M_{i_1, \dots, i_m} \quad (10.4)$$

for all collections  $i_1 < i_2 < \cdots < i_m$  of the numbers  $1, \dots, n$ . Thus we assign to the subspace  $M$  as many of the numbers  $p_{i_1, \dots, i_m}$  as there are combinations of  $n$  things taken  $m$  at a time, that is,  $\nu = C_n^m$ . From formula (10.3) and the condition that the rank of the matrix  $M$  is equal to  $m$ , it follows that these numbers  $p_{i_1, \dots, i_m}$  cannot all become zero simultaneously. On the other hand, formula (10.2) shows that in replacing the basis  $\mathbf{a}_1, \dots, \mathbf{a}_m$  of the subspace  $M$  by some other basis  $\mathbf{b}_1, \dots, \mathbf{b}_m$  of this subspace, all these numbers are simultaneously multiplied by some number  $c \neq 0$ . Thus the numbers  $p_{i_1, \dots, i_m}$  for  $i_1 < i_2 < \cdots < i_m$  can be taken as the homogeneous coordinates of a point of the projective space  $\mathbb{P}^{\nu-1} = \mathbb{P}(N)$ , where  $\dim N = \nu$  and  $\dim \mathbb{P}(N) = \nu - 1$ .

**Definition 10.1** The totality of numbers  $p_{i_1, \dots, i_m}$  in (10.4) for all collections  $i_1 < i_2 < \cdots < i_m$  taking the values  $1, \dots, n$  is called the *Plücker coordinates* of the  $m$ -dimensional subspace  $M \subset L$ .

As we have seen, Plücker coordinates are defined only up to a common nonzero factor; the collection of them must be understood as a point in the projective space  $\mathbb{P}^{\nu-1}$ .

The simplest special case  $m = 1$  returns us to the definition of projective space, whose points correspond to one-dimensional subspaces  $\langle \mathbf{a} \rangle$  of some vector space  $L$ . The numbers  $p_{i_1, \dots, i_m}$  in this case become the homogeneous coordinates of a point. It is therefore not surprising that all of these depend on the choice of a coordinate system (that is, a basis) of the space  $L$ . Following tradition, in the sequel we shall allow for a certain imprecision and call “Plücker coordinates” of the subspace  $M$  both a point of the projective space  $\mathbb{P}^{\nu-1}$  and the collection of numbers  $p_{i_1, \dots, i_m}$  specified in this definition.

**Theorem 10.2** *The Plücker coordinates of a subspace  $M \subset L$  uniquely determine the subspace.*

*Proof* Let us choose an arbitrary basis  $\mathbf{a}_1, \dots, \mathbf{a}_m$  of the subspace  $M$ . It uniquely determines (and not up to a common factor) the minors  $M_{i_1, \dots, i_m}$ , without regard to the order of the indices  $i_1, \dots, i_m$ . The minors are uniquely determined by the Plücker coordinates (10.4), according to formula (10.3).

A vector  $\mathbf{x} \in L$  belongs to the subspace  $M = \langle \mathbf{a}_1, \dots, \mathbf{a}_m \rangle$  if and only if the rank of the matrix

$$\overline{M} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \\ x_1 & x_2 & \cdots & x_n \end{pmatrix},$$

consisting of the coordinates of the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m, \mathbf{x}$  in some (arbitrary) basis of the space  $L$ , is equal to  $m$ , that is, if all the minors of order  $m + 1$  of the matrix  $\overline{M}$  are equal to zero. Let us consider the minor that comprises the columns with indices forming the subset  $X = \{k_1, \dots, k_{m+1}\}$  of the set  $\mathbb{N}_n = \{1, \dots, n\}$ , where we may

assume that  $k_1 < k_2 < \dots < k_{m+1}$ . Expanding it along the last row, we obtain the equality

$$\sum_{\alpha \in X} x_\alpha A_\alpha = 0, \quad (10.5)$$

where  $A_\alpha$  is the cofactor of the element  $x_\alpha$  in the minor under consideration. But by definition, the minor corresponding to  $A_\alpha$  is obtained from the matrix  $\overline{M}$  by deleting the last row and the column with index  $\alpha$ . Therefore, it coincides with one of the minors of the matrix  $M$ , and the indices of its columns are obtained by deleting the element  $\alpha$  from the set  $X$ . For writing the sets thus obtained, one frequently uses the convenient notation

$$\{k_1, \dots, \check{k}_\alpha, \dots, k_{m+1}\},$$

where the notation  $\check{\phantom{x}}$  signifies the omission of the element so indicated. Thus relationship (10.5) can be written in the form

$$\sum_{j=1}^{m+1} (-1)^j x_{k_j} M_{k_1, \dots, \check{k}_j, \dots, k_{m+1}} = 0. \quad (10.6)$$

Since the minors  $M_{i_1, \dots, i_m}$  of the matrix  $M$  are expressed in Plücker coordinates by formula (10.4), relationships (10.6), obtained from all possible subsets  $X = \{k_1, \dots, k_{m+1}\}$  of the set  $\mathbb{N}_n$ , also give expressions in terms of Plücker coordinates of the condition  $\mathbf{x} \in M$ , which completes the proof of the theorem.  $\square$

By Theorem 10.2, Plücker coordinates uniquely define the subspace  $M$ , but as a rule, they cannot assume arbitrary values. It is true that for  $m = 1$ , the homogeneous coordinates of a point of projective space can be chosen with arbitrary numbers (of course, with the exception of the one collection consisting of all zeros). Another equally simple case is  $m = n - 1$ , in which subspaces are hyperplanes corresponding to points of  $\mathbb{P}(L^*)$ . Hyperplanes are defined by their coordinates in this projective space, which also can be chosen as arbitrary collections of numbers (again with the exclusion of the collection consisting of all zeros). It is not difficult to verify that these homogeneous coordinates can differ from Plücker coordinates only by their signs, that is, by the factor  $\pm 1$ . However, as we shall now see, for an arbitrary number  $m < n$ , the Plücker coordinates are connected to one another by certain specific relationships.

*Example 10.3* Let us consider the next case in order of complexity:  $n = 4$ ,  $m = 2$ . If we pass to projective spaces corresponding to  $L$  and  $M$ , then this will give us a description of the totality of projective lines in three-dimensional projective space (the case considered by Plücker).

Since  $n = 4$ ,  $m = 2$ , we have  $\nu = C_4^2 = 6$ , and consequently, each plane  $M \subset L$  has six Plücker coordinates:

$$p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34}. \quad (10.7)$$

It is easy to see that for an arbitrary basis of the space  $L$ , we may always choose a basis  $\mathbf{a}, \mathbf{b}$  in the subspace  $M$  in such a way that the matrix  $M$  given by formula (10.1) will have the form

$$M = \begin{pmatrix} 1 & 0 & \alpha & \beta \\ 0 & 1 & \gamma & \delta \end{pmatrix}.$$

From this follow easily the values of the Plücker coordinates (10.7):

$$\begin{aligned} p_{12} &= 1, & p_{13} &= \gamma, & p_{14} &= \delta, & p_{23} &= -\alpha, & p_{24} &= -\beta, \\ p_{34} &= \alpha\delta - \beta\gamma, \end{aligned}$$

which yields the relationship  $p_{34} - p_{13}p_{24} + p_{14}p_{23} = 0$ . In order to make this homogeneous, we will use the fact that  $p_{12} = 1$ , and write it in the form

$$p_{12}p_{34} - p_{13}p_{24} + p_{14}p_{23} = 0. \quad (10.8)$$

The relationship (10.8) is already homogeneous, and therefore, it is preserved under multiplication of all the Plücker coordinates (10.7) by an arbitrary nonzero factor  $c$ . Thus relationship (10.8) remains valid for an arbitrary choice of Plücker coordinates, and this means that it defines a point in some projective algebraic variety in 5-dimensional projective space.<sup>1</sup> In the following section, we shall study an analogous question in the general case, for arbitrary dimension  $m < n$ .

## 10.2 The Plücker Relations and the Grassmannian

We shall now describe the relationships satisfied by Plücker coordinates of an  $m$ -dimensional subspace  $M$  of an  $n$ -dimensional space  $L$  for arbitrary  $n$  and  $m$ . Here we shall use the following notation and conventions. Although in the definition of Plücker coordinates  $p_{i_1, \dots, i_m}$  it was assumed that  $i_1 < i_2 < \dots < i_m$ , now we shall consider numbers  $p_{i_1, \dots, i_m}$  also with other collections of indices. Namely, if  $(j_1, \dots, j_m)$  is an arbitrary collection of  $m$  indices taking the values  $1, \dots, n$ , then we set

$$p_{j_1, \dots, j_m} = 0 \quad (10.9)$$

if some two of the numbers  $j_1, \dots, j_m$  are equal, while if all the numbers  $j_1, \dots, j_m$  are distinct and  $(i_1, \dots, i_m)$  is their arrangement in ascending order, then we set

$$p_{j_1, \dots, j_m} = \pm p_{i_1, \dots, i_m}, \quad (10.10)$$

---

<sup>1</sup>This variety is called a *quadric*.

where the sign  $+$  or  $-$  depends on whether the permutation that takes  $(j_1, \dots, j_m)$  to  $(i_1, \dots, i_m)$  is even or odd (that is, whether the number of transpositions is even or odd), according to Theorem 2.25.

In other words, in view of equality (10.3), let us set

$$p_{j_1, \dots, j_m} = M_{j_1, \dots, j_m}, \quad (10.11)$$

where  $(j_1, \dots, j_m)$  is an arbitrary collection of indices assuming the values  $1, \dots, n$ .

**Theorem 10.4** *For every  $m$ -dimensional subspace  $M$  of an  $n$ -dimensional space  $L$  and for any two sets  $(j_1, \dots, j_{m-1})$  and  $(k_1, \dots, k_{m+1})$  of indices taking the values  $1, \dots, n$ , the following relationships hold:*

$$\sum_{r=1}^{m+1} (-1)^r p_{j_1, \dots, j_{m-1}, k_r} \cdot p_{k_1, \dots, \check{k}_r, \dots, k_{m+1}} = 0. \quad (10.12)$$

*These are called the Plücker relations.*

The notation  $k_1, \dots, \check{k}_r, \dots, k_{m+1}$  means that we omit  $k_r$  in the sequence  $k_1, \dots, k_r, \dots, k_{m+1}$ .

Let us note that the indices among the numbers  $p_{\alpha_1, \dots, \alpha_m}$  entering relationship (10.12) are not necessarily in ascending order, so they are not Plücker coordinates. But with the aid of relationships (10.9) and (10.10), we can easily express them in terms of Plücker coordinates. Therefore, relationship (10.12) may also be viewed as a relationship among Plücker coordinates.

*Proof of Theorem 10.4* Returning to the definition of Plücker coordinates in terms of the minors of the matrix (10.1) and using relationship (10.11), we see that equality (10.12) can be rewritten in the form

$$\sum_{r=1}^{m+1} (-1)^r M_{j_1, \dots, j_{m-1}, k_r} \cdot M_{k_1, \dots, \check{k}_r, \dots, k_{m+1}} = 0. \quad (10.13)$$

Let us show that relationship (10.13) holds for the minors of an arbitrary matrix of type  $(m, n)$ . To this end, let us expand the determinant  $M_{j_1, \dots, j_{m-1}, k_r}$  along the last column. Let us denote the cofactor of the element  $a_{lk_r}$  of the last column of this determinant by  $A_l$ ,  $l = 1, \dots, m$ . Thus the cofactor  $A_l$  corresponds to the minor located in the rows and columns with indices  $(1, \dots, \check{l}, \dots, m)$  and  $(j_1, \dots, j_{m-1})$  respectively. Then

$$M_{j_1, \dots, j_{m-1}, k_r} = \sum_{l=1}^m a_{lk_r} A_l.$$



On substituting this expression into the left-hand side of relationship (10.13), we arrive at the equality

$$\begin{aligned} & \sum_{r=1}^{m+1} (-1)^r M_{j_1, \dots, j_{m-1}, k_r} \cdot M_{k_1, \dots, \check{k}_r, \dots, k_{m+1}} \\ &= \sum_{r=1}^{m+1} (-1)^r \left( \sum_{l=1}^m a_{lk_r} A_l \right) M_{k_1, \dots, \check{k}_r, \dots, k_{m+1}}. \end{aligned}$$

Changing the order of summation, we obtain

$$\begin{aligned} & \sum_{r=1}^{m+1} (-1)^r M_{j_1, \dots, j_{m-1}, k_r} \cdot M_{k_1, \dots, \check{k}_r, \dots, k_{m+1}} \\ &= \sum_{l=1}^m \left( \sum_{r=1}^{m+1} (-1)^r a_{lk_r} M_{k_1, \dots, \check{k}_r, \dots, k_{m+1}} \right) A_l. \end{aligned}$$

But the sum in parentheses is equal to the result of the expansion along the first row of the determinant of the square matrix of order  $m+1$  consisting of the columns of the matrix (10.1) numbered  $k_1, \dots, k_{m+1}$  and rows numbered  $l, 1, \dots, m$ . This determinant is equal to

$$\begin{vmatrix} a_{lk_1} & a_{lk_2} & \cdots & a_{lk_{m+1}} \\ a_{1k_1} & a_{1k_2} & \cdots & a_{1k_{m+1}} \\ a_{2k_1} & a_{2k_2} & \cdots & a_{2k_{m+1}} \\ \vdots & \vdots & \ddots & \vdots \\ a_{mk_1} & a_{mk_2} & \cdots & a_{mk_{m+1}} \end{vmatrix} = 0.$$

Indeed, for arbitrary  $l = 1, \dots, m$ , two of its rows (numbered 1 and  $l+1$ ) coincide, and this means that the determinant is equal to zero.  $\square$

*Example 10.5* Let us return once more to the case  $n = 4$ ,  $m = 2$  considered in the previous section. Relationships (10.12) are here determined by subsets  $(k)$  and  $(l, m, n)$  of the set  $\{1, 2, 3, 4\}$ . If, for example,  $k = 1$  and  $l = 2$ ,  $m = 3$ ,  $n = 4$ , then we obtain relationship (10.8) introduced earlier. It is easily verified that if all the numbers  $k, l, m, n$  are distinct, then we obtain the same relationship (10.8), while if among them there are two that are equal, then relationship (10.12) is an identity (for the proof of this, we can use the antisymmetry of  $p_{ij}$  with respect to  $i$  and  $j$ ). Therefore, in the general case, too (for arbitrary  $m$  and  $n$ ), relationships (10.12) among the Plücker coordinates are called the *Plücker relations*.

We have seen that to each subspace  $M$  of given dimension  $m$  of the space  $L$  of dimension  $n$ , there correspond its Plücker coordinates

$$p_{i_1, \dots, i_m}, \quad i_1 < i_2 < \cdots < i_m, \quad (10.14)$$

satisfying the relationships (10.12). Thus an  $m$ -dimensional subspace  $M \subset L$  is determined by its Plücker coordinates (10.14), completely analogously to how points of a projective space are determined by their homogeneous coordinates (this is in fact a special case of Plücker coordinates for  $m = 1$ ). However, for  $m > 1$ , the coordinates of the subspace  $M$  cannot be assigned arbitrarily: it is necessary that they satisfy relationships (10.12). Below, we shall prove that these relationships are also sufficient for the collection of numbers (10.14) to be Plücker coordinates of some  $m$ -dimensional subspace  $M \subset L$ . For this, we shall find the following geometric interpretation of Plücker coordinates useful.

Relationships (10.12) are homogeneous (of degree 2) with respect to the numbers  $p_{i_1, \dots, i_m}$ . After substitution on the basis of formulas (10.9) and (10.10), each of these relationships remains homogeneous, and thus they define a certain projective algebraic variety in the projective space  $\mathbb{P}^{v-1}$ , called a *Grassmann variety* or simply *Grassmannian* and denoted by  $G(m, n)$ .

We shall now investigate the Grassmannian  $G(m, n)$  in greater detail.

As we have seen,  $G(m, n)$  is contained in the projective space  $\mathbb{P}^{v-1}$ , where  $v = C_n^m$  (see p. 351), and the homogeneous coordinates are written as the numbers (10.14) with all possible increasing collections of indices taking the values  $1, \dots, n$ . The space  $\mathbb{P}^{v-1}$  is the union of affine subsets  $U_{i_1, \dots, i_m}$ , each of which is defined by the condition  $p_{i_1, \dots, i_m} \neq 0$  for some choice of indices  $i_1, \dots, i_m$ . From this we obtain

$$G(m, n) = \bigcup_{i_1, \dots, i_m} (G(m, n) \cap U_{i_1, \dots, i_m}).$$

We shall investigate separately one of these subsets  $G(m, n) \cap U_{i_1, \dots, i_m}$ , for example, for simplicity, the subset with indices  $(i_1, \dots, i_m) = (1, \dots, m)$ . The general case is considered completely analogously and differs only in the numeration of the coordinates in the space  $\mathbb{P}^{v-1}$ . We may assume that for points of our affine subset  $U_{1, \dots, m}$ , the number  $p_{1, \dots, m}$  is equal to 1.

Relationships (10.12) give the possibility to choose Plücker coordinates (10.14) of the subspace  $M$  (or equivalently, the minors  $M_{i_1, \dots, i_m}$  of the matrix (10.1)) in the form of polynomials in coordinates  $p_{i_1, \dots, i_m}$ , such that among the indices  $i_1 < i_2 < \dots < i_m$ , not more than one exceeds  $m$ . Any such collection of indices obviously has the form  $(1, \dots, \check{r}, \dots, m, l)$ , where  $r \leq m$  and  $l > m$ . Let us denote the Plücker coordinate corresponding to this collection by  $\overline{p}_{rl}$ , that is, we set  $\overline{p}_{rl} = p_{1, \dots, \check{r}, \dots, m, l}$ .

Let us consider an arbitrary ordered collection  $j_1 < j_2 < \dots < j_m$  of numbers between 1 and  $n$ . If the indices  $j_k$  are less than or equal to  $m$  for all  $k = 1, \dots, m$ , then the collection  $(j_1, j_2, \dots, j_m)$  coincides with the collection  $(1, 2, \dots, m)$ , and since the Plücker coordinate  $p_{1, \dots, m}$  is equal to 1, there is nothing to prove. Thus we have only to consider the remaining case.

Let  $j_k > m$  be one of the numbers  $j_1 < j_2 < \dots < j_m$ . Let us use relationship (10.12), corresponding to the collection  $(j_1, \dots, \check{j}_k, \dots, j_m)$  of  $m - 1$  numbers and the collection  $(1, \dots, m, j_k)$  of  $m + 1$  numbers. In this case, relationship (10.12)

assumes the form

$$\sum_{r=1}^m (-1)^r p_{j_1, \dots, \check{j}_k, \dots, j_m, r} \cdot p_{1, 2, \dots, \check{r}, \dots, m, j_k} + (-1)^{m+1} p_{j_1, \dots, \check{j}_k, \dots, j_m, j_k} = 0,$$

since  $p_{1, \dots, m} = 1$ . In view of the antisymmetry of the expression  $p_{j_1, \dots, j_m}$ , it follows that  $p_{j_1, \dots, j_m} = p_{j_1, \dots, \check{j}_k, \dots, j_m, j_k}$  is equal to the sum (with alternating signs) of the products  $p_{j_1, \dots, \check{j}_k, \dots, j_m, r} \bar{p}_{r, l}$ . If among the numbers  $j_1, \dots, j_m$  there were  $s$  numbers exceeding  $m$ , then among the numbers  $j_1, \dots, \check{j}_k, \dots, j_m$ , there would be already  $s - 1$  of them.

Repeating this process as many times as necessary, we will obtain as a result an expression of the chosen Plücker coordinate  $p_{j_1, \dots, j_m}$  in terms of the coordinates  $\bar{p}_{r, l}$ ,  $r \leq m$ ,  $l > m$ . We have thereby obtained the following important result.

**Theorem 10.6** *For each point in the set  $G(m, n) \cap U_{1, \dots, m}$ , all the Plücker coordinates (10.14) are polynomials in the coordinates  $\bar{p}_{r, l} = p_{1, \dots, \check{r}, \dots, m, l}$ ,  $r \leq m$ ,  $l > m$ .*

Since the numbers  $r$  and  $l$  satisfy  $1 \leq r \leq m$  and  $m < l \leq n$ , it follows that all possible collections of coordinates  $\bar{p}_{r, l}$  form an affine subspace  $V$  of dimension  $m(n - m)$ . By Theorem 10.6, all the remaining Plücker coordinates  $p_{i_1, \dots, i_m}$  are polynomials in  $\bar{p}_{r, l}$ , and therefore the coordinates  $\bar{p}_{r, l}$  uniquely define a point of the set  $G(m, n) \cap U_{1, \dots, m}$ . Thus is obtained a natural bijection (given by these polynomials) between points of the set  $G(m, n) \cap U_{1, \dots, m}$  and points of the affine space  $V$  of dimension  $m(n - m)$ . Of course, the same is true as well for points of any other set  $G(m, n) \cap U_{i_1, \dots, i_m}$ . In algebraic geometry, this fact is expressed by saying that the Grassmannian  $G(m, n)$  is covered by the affine space of dimension  $m(n - m)$ .

**Theorem 10.7** *Every point of the Grassmannian  $G(m, n)$  corresponds to some  $m$ -dimensional subspace  $M \subset L$  as described in the previous section.*

*Proof* Since the Grassmannian  $G(m, n)$  is the union of sets  $G(m, n) \cap U_{i_1, \dots, i_m}$ , it suffices to prove the theorem for each set separately. We shall carry out the proof for the set  $G(m, n) \cap U_{1, \dots, m}$ , since the rest differ from it only in the numeration of coordinates.

Let us choose an  $m$ -dimensional subspace  $M \subset L$  and basis  $\mathbf{a}_1, \dots, \mathbf{a}_m$  in it so that in the associated matrix  $M$  given by formula (10.1), the elements residing in its first  $m$  columns take the form of the identity matrix  $E$  of order  $m$ . Then the matrix  $M$  has the form

$$M = \begin{pmatrix} 1 & 0 & \cdots & 0 & a_{1m+1} & \cdots & a_{1n} \\ 0 & 1 & \cdots & 0 & a_{2m+1} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & a_{mm+1} & \cdots & a_{mn} \end{pmatrix}. \quad (10.15)$$

By Theorem 10.6, the Plücker coordinates (10.14) are polynomials in  $\bar{p}_{r, l} = p_{1, \dots, \check{r}, \dots, m, l}$ . Moreover, by the definition of Plücker coordinates (10.4), we have

$p_{1,\dots,\check{r},\dots,m,l} = M_{1,\dots,\check{r},\dots,m,l}$ . Here, in the  $r$ th row of the minor  $M_{1,\dots,\check{r},\dots,m,l}$  of the matrix (10.15), all elements are equal to zero, except for the element in the last ( $l$ )th column, which is equal to  $a_{rl}$ . Expanding the minor  $M_{1,\dots,\check{r},\dots,m,l}$  along the  $r$ th row, we see that it is equal to  $(-1)^{r+l}a_{rl}$ . In other words,  $\bar{p}_{rl} = (-1)^{r+l}a_{rl}$ .

By our construction, all elements  $a_{rl}$  of the matrix (10.15) can assume arbitrary values by the choice of a suitable subspace  $M \subset L$  and basis  $\mathbf{a}_1, \dots, \mathbf{a}_m$  in it. Thus the Plücker coordinates  $\bar{p}_{rl}$  also assume arbitrary values. It remains to observe that by Theorem 10.6, all remaining Plücker coordinates are polynomials in  $\bar{p}_{rl}$ , and consequently, for the constructed subspace  $M$ , they determine the given point of the set  $G(m, n) \cap U_{1,\dots,m}$ .  $\square$

### 10.3 The Exterior Product

Now we shall attempt to understand the sense in which the subspace  $M \subset L$  is related to its Plücker coordinates, after separating out those parts of the construction that depend on the choice of bases  $\mathbf{e}_1, \dots, \mathbf{e}_n$  in  $L$  and  $\mathbf{a}_1, \dots, \mathbf{a}_m$  in  $M$  from those that do not depend on the choice of basis.

Our definition of Plücker coordinates was connected with the minors of the matrix  $M$  given by formula (10.1), and since minors (like all determinants) are multilinear and antisymmetric functions of the rows (and columns), let us begin by recalling the appropriate definitions from Sect. 2.6 (especially because now we shall need them in a somewhat changed form). Namely, while in Chap. 2, we considered only functions of rows, now we shall consider functions of vectors belonging to an arbitrary vector space  $L$ . We shall assume that the space  $L$  is finite-dimensional. Then by Theorem 3.64, it is isomorphic to the space of rows of length  $n = \dim L$ , and so we might have used the definitions from Sect. 2.6. But such an isomorphism itself depends on the choice of basis in the space  $L$ , and our goal is precisely to study the dependence of our construction on the choice of basis.

**Definition 10.8** A function  $F(\mathbf{x}_1, \dots, \mathbf{x}_m)$  in  $m$  vectors of the space  $L$  taking numeric values is said to be *multilinear* if for every index  $i$  in the range 1 to  $m$  and arbitrary fixed vectors  $\mathbf{a}_1, \dots, \check{\mathbf{a}}_i, \dots, \mathbf{a}_m$ ,

$$F(\mathbf{a}_1, \dots, \mathbf{a}_{i-1}, \mathbf{x}_i, \mathbf{a}_{i+1}, \dots, \mathbf{a}_m)$$

is a linear function of the vector  $\mathbf{x}_i$ .

For  $m = 1$ , we arrive at the notion of linear function introduced in Sect. 3.7, and for  $m = 2$ , this is the notion of bilinear form, introduced in Sect. 6.1.

The definition of antisymmetric function given in Sect. 2.6 was valid for every set, and in particular, we may apply it to the set of all vectors of the space  $L$ . According to this definition, for every pair of distinct indices  $r$  and  $s$  in the range 1 to  $m$ , the relationship

$$F(\mathbf{x}_1, \dots, \mathbf{x}_r, \dots, \mathbf{x}_s, \dots, \mathbf{x}_m) = -F(\mathbf{x}_1, \dots, \mathbf{x}_s, \dots, \mathbf{x}_r, \dots, \mathbf{x}_m) \quad (10.16)$$

must be satisfied for every collection of vectors  $\mathbf{x}_1, \dots, \mathbf{x}_m \in L$ . As proved in Sect. 2.6, it suffices to prove property (10.16) for  $s = r + 1$ , that is, a transposition of two neighboring vectors from the collection  $\mathbf{x}_1, \dots, \mathbf{x}_m$  is performed. Then property (10.16) will also be satisfied for arbitrary indices  $r$  and  $s$ . In view of this, we shall often formulate the condition of antisymmetry only for “neighboring” indices and use the fact that it then holds for two arbitrary indices  $r$  and  $s$ .

If these numbers are elements of a field of characteristic different from 2, then it follows that  $F(\mathbf{x}_1, \dots, \mathbf{x}_m) = 0$  if any two vectors  $\mathbf{x}_1, \dots, \mathbf{x}_m$  coincide.

Let us denote by  $\Pi^m(L)$  the collection of all multilinear functions of  $m$  vectors of the space  $L$ , and by  $\Omega^m(L)$  the collection of all antisymmetric functions in  $\Pi^m(L)$ . The sets  $\Pi^m(L)$  and  $\Omega^m(L)$  become vector spaces if for all  $F, G \in \Pi^m(L)$  we define their sum  $H = F + G \in \Pi^m(L)$  by the formula

$$H(\mathbf{x}_1, \dots, \mathbf{x}_m) = F(\mathbf{x}_1, \dots, \mathbf{x}_m) + G(\mathbf{x}_1, \dots, \mathbf{x}_m)$$

and define for every function  $F \in \Pi^m(L)$  the product by the scalar  $\alpha$  as the function  $H = \alpha F \in \Pi^m(L)$  according to the formula

$$H(\mathbf{x}_1, \dots, \mathbf{x}_m) = \alpha F(\mathbf{x}_1, \dots, \mathbf{x}_m).$$

It directly follows from these definitions that  $\Pi^m(L)$  is thereby converted to a vector space, and  $\Omega^m(L) \subset \Pi^m(L)$  is a subspace of  $\Pi^m(L)$ .

Let  $\dim L = n$ , and let  $\mathbf{e}_1, \dots, \mathbf{e}_n$  be some basis of the space  $L$ . It follows from the definition that the multilinear function  $F(\mathbf{x}_1, \dots, \mathbf{x}_m)$  is defined for all collections of vectors  $(\mathbf{x}_1, \dots, \mathbf{x}_m)$  if it is defined for those collections whose vectors  $\mathbf{x}_i$  belong to our basis. Indeed, repeating the arguments from Sect. 2.7 verbatim that we used in the proof of Theorem 2.29, we obtain for  $F(\mathbf{x}_1, \dots, \mathbf{x}_m)$  the same formulas (2.40) and (2.43). Thus for the chosen basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , the multilinear function  $F(\mathbf{x}_1, \dots, \mathbf{x}_m)$  is determined by its values  $F(\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_m})$ , where  $i_1, \dots, i_m$  are all possible collections of numbers from the set  $\mathbb{N}_n = \{1, \dots, n\}$ .

The previous line of reasoning shows that the space  $\Pi^m(L)$  is isomorphic to the space of functions on the set  $\mathbb{N}_n^m = \mathbb{N}_n \times \dots \times \mathbb{N}_n$  ( $m$ -fold product). It follows that the dimension of the space  $\Pi^m(L)$  is finite and coincides with the number of elements of the set  $\mathbb{N}_n^m$ . It is easy to verify that this number is equal to  $n^m$ , and so  $\dim \Pi^m(L) = n^m$ .

As we observed in Example 3.36 (p. 94), in a space of functions  $f$  on a finite set  $\mathbb{N}_n^m$ , there exists a basis consisting of  $\delta$ -functions assuming the value 1 on one element of  $\mathbb{N}_n^m$  and the value 0 on all the other elements (p. 94). In our case, we shall introduce a special notation for such a basis. Let  $\mathbf{I} = (i_1, \dots, i_m)$  be an arbitrary element of the set  $\mathbb{N}_n^m$ . Then we denote by  $f_{\mathbf{I}}$  the function taking the value 1 at the element  $\mathbf{I}$  and the value 0 on all remaining elements of the set  $\mathbb{N}_n^m$ .

We now move on to an examination of the subspace of antisymmetric multilinear functions  $\Omega^m(L)$ , assuming as previously that there has been chosen in  $L$  some basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$ . To verify that a multilinear function  $F$  is antisymmetric, it is necessary and sufficient that property (10.16) be satisfied for the vectors  $\mathbf{e}_i$  of the basis. In

other words, this reduces to the relationships

$$F(\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_r}, \dots, \mathbf{e}_{i_s}, \dots, \mathbf{e}_{i_m}) = -F(\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_s}, \dots, \mathbf{e}_{i_r}, \dots, \mathbf{e}_{i_m})$$

for all collections of vectors  $\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_m}$  in the chosen basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the space  $L$ . Therefore, for every function  $F \in \Omega^m(L)$  and every collection  $(j_1, \dots, j_m) \in \mathbb{N}_n^m$ , we have the equality

$$F(\mathbf{e}_{j_1}, \dots, \mathbf{e}_{j_m}) = \pm F(\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_m}), \quad (10.17)$$

where the numbers  $i_1, \dots, i_m$  are the same as  $j_1, \dots, j_m$ , but arranged in ascending order  $i_1 < i_2 < \dots < i_m$ , while the sign  $+$  or  $-$  in (10.17) depends on whether the number of transpositions necessary for passing from the collection  $(i_1, \dots, i_m)$  to the collection  $(j_1, \dots, j_m)$  is even or odd (we note that if any two of the numbers  $j_1, \dots, j_m$  are equal, then both sides of equality (10.17) become equal to zero).

Reasoning just as in the case of the space  $\Pi^m(L)$ , we conclude that the space  $\Omega^m(L)$  is isomorphic to the space of functions on the set  $\vec{\mathbb{N}}_n^m \subset \mathbb{N}_n^m$ , which consists of all *increasing* sets  $\mathbf{I} = (i_1, \dots, i_m)$ , that is, those for which  $i_1 < i_2 < \dots < i_m$ . From this it follows in particular that  $\Omega^m(L) = \{0\}$  if  $m > n$ . It is easy to see that the number of such increasing sets  $\mathbf{I}$  is equal to  $C_n^m$ , and therefore,

$$\dim \Omega^m(L) = C_n^m. \quad (10.18)$$

We shall denote by  $F_{\mathbf{I}}$  the  $\delta$ -function of the space  $\Omega^m(L)$ , taking the value 1 on the set  $\mathbf{I} \in \vec{\mathbb{N}}_n^m$  and the value 0 on all the remaining sets in  $\vec{\mathbb{N}}_n^m$ .

The vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m \in L$  determine on the space  $\Omega^m(L)$  a linear function  $\varphi$  given by the relationship

$$\varphi(F) = F(\mathbf{a}_1, \dots, \mathbf{a}_m) \quad (10.19)$$

for an arbitrary element  $F \in \Omega^m(L)$ . Thus  $\varphi$  is a linear function on  $\Omega^m(L)$ , that is, an element of the dual space  $\Omega^m(L)^*$ .

**Definition 10.9** The dual space  $\Lambda^m(L) = \Omega^m(L)^*$  is called the *space of  $m$ -vectors* or the  *$m$ th exterior power* of the space  $L$ , and its elements are called  *$m$ -vectors*. A vector  $\varphi \in \Lambda^m(L)$  constructed with the help of relationship (10.19) involving the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m$  is called the *exterior product* (or *wedge product*) of  $\mathbf{a}_1, \dots, \mathbf{a}_m$  and is denoted by

$$\varphi = \mathbf{a}_1 \wedge \mathbf{a}_2 \wedge \dots \wedge \mathbf{a}_m.$$

Now let us explore the connection between the exterior product and Plücker coordinates of the subspace  $M \subset L$ . To this end, it is necessary to choose some basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  in  $L$  and some basis  $\mathbf{a}_1, \dots, \mathbf{a}_m$  in  $M$ . The Plücker coordinates of the subspace  $M$  take the form (10.4), where  $M_{i_1, \dots, i_m}$  is the minor of the matrix (10.1) that resides in columns  $i_1, \dots, i_m$  and is an antisymmetric function of its columns. Let us introduce for the Plücker coordinates and associated minors the notation

$$p_{\mathbf{I}} = p_{i_1, \dots, i_m}, \quad M_{\mathbf{I}} = M_{i_1, \dots, i_m}, \quad \text{where } \mathbf{I} = (i_1, \dots, i_m) \in \vec{\mathbb{N}}_n^m.$$

To the basis of the space  $\Omega^m(\mathbf{L})$  consisting of  $\delta$ -functions  $F_I$ , there corresponds the dual basis, of the dual space  $\Lambda^m(\mathbf{L})$ , whose vectors we shall denote by  $\varphi_I$ . Using the notation that we introduced in Sect. 3.7, we may say that the dual basis is defined by the condition

$$(F_I, \varphi_I) = 1 \quad \text{for all } I \in \vec{\mathbb{N}}_n^m, \quad (F_I, \varphi_J) = 0 \quad \text{for all } I \neq J. \quad (10.20)$$

In particular, the vector  $\varphi = \mathbf{a}_1 \wedge \mathbf{a}_2 \wedge \cdots \wedge \mathbf{a}_m$  of the space  $\Lambda^m(\mathbf{L})$  can be expressed as a linear combination of vectors in this basis:

$$\varphi = \sum_{I \in \vec{\mathbb{N}}_n^m} \lambda_I \varphi_I \quad (10.21)$$

with certain coefficients  $\lambda_I$ . Using formulas (10.19) and (10.20), we obtain the following equality:

$$\lambda_I = \varphi(F_I) = F_I(\mathbf{a}_1, \dots, \mathbf{a}_m).$$

For determining the values  $F_I(\mathbf{a}_1, \dots, \mathbf{a}_m)$ , we may make use of Theorem 2.29; see formulas (2.40) and (2.43). Since  $F_I(\mathbf{e}_{j_1}, \dots, \mathbf{e}_{j_m}) = 0$  when the indices of  $\mathbf{e}_{j_1}, \dots, \mathbf{e}_{j_m}$  form the collection  $J \neq I$ , then from formula (2.43), it follows that the values  $F_I(\mathbf{a}_1, \dots, \mathbf{a}_m)$  depend only on the elements appearing in the minor  $M_I$ . The minor  $M_I$  is a linear and antisymmetric function of its rows. In view of the fact that by definition,  $F_I(\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_m}) = 1$ , we obtain from Theorem 2.15 that  $F_I(\mathbf{a}_1, \dots, \mathbf{a}_m) = M_I = p_I$ . In other words, we have the equality

$$\varphi = \mathbf{a}_1 \wedge \mathbf{a}_2 \wedge \cdots \wedge \mathbf{a}_m = \sum_{I \in \vec{\mathbb{N}}_n^m} M_I \varphi_I = \sum_{I \in \vec{\mathbb{N}}_n^m} p_I \varphi_I. \quad (10.22)$$

Thus any collection of  $m$  vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m$  uniquely determines the vector  $\mathbf{a}_1 \wedge \cdots \wedge \mathbf{a}_m$  in the space  $\Lambda^m(\mathbf{L})$ , where the Plücker coordinates of the subspace  $\langle \mathbf{a}_1, \dots, \mathbf{a}_m \rangle$  are the coordinates of this vector  $\mathbf{a}_1 \wedge \cdots \wedge \mathbf{a}_m$  with respect to the basis  $\varphi_I$ ,  $I \in \vec{\mathbb{N}}_n^m$ , of the space  $\Lambda^m(\mathbf{L})$ . Like all coordinates, they depend on this basis, which itself is constructed as the dual basis to some basis of the space  $\Omega^m(\mathbf{L})$ .

**Definition 10.10** A vector  $\mathbf{x} \in \Lambda^m(\mathbf{L})$  is said to be *decomposable* if it can be represented as an exterior product

$$\mathbf{x} = \mathbf{a}_1 \wedge \mathbf{a}_2 \wedge \cdots \wedge \mathbf{a}_m \quad (10.23)$$

with some  $\mathbf{a}_1, \dots, \mathbf{a}_m \in \mathbf{L}$ .

Let the  $m$ -vector  $\mathbf{x}$  have coordinates  $x_{i_1, \dots, i_m}$  in some basis  $\varphi_I$ ,  $I \in \vec{\mathbb{N}}_n^m$ , of the space  $\Lambda^m(\mathbf{L})$ . As in the case of an arbitrary vector space, the coordinates  $x_{i_1, \dots, i_m}$  can assume arbitrary values in the associated field. In order for an  $m$ -vector  $\mathbf{x}$  to be decomposable, that is, that it satisfy the relationship (10.23) with some vectors

$\mathbf{a}_1, \dots, \mathbf{a}_m \in L$ , it is necessary and sufficient that its coordinates  $x_{i_1, \dots, i_m}$  coincide with the Plücker coordinates  $p_{i_1, \dots, i_m}$  of the subspace  $M = \langle \mathbf{a}_1, \dots, \mathbf{a}_m \rangle$  in  $L$ . But as we established in the previous section, the collection of Plücker coordinates of a subspace  $M \subset L$  cannot be an arbitrary collection of  $\nu$  numbers, but only one that satisfies the Plücker relations (10.12). Consequently, the Plücker relations give necessary and sufficient conditions for an  $m$ -vector  $\mathbf{x}$  to be decomposable.

Thus for the specification of  $m$ -dimensional subspaces  $M \subset L$ , we need only the decomposable  $m$ -vectors (the indecomposable  $m$ -vectors correspond to no  $m$ -dimensional subspace). However, generally speaking, the decomposable vectors do not form a vector space (the sum of two decomposable vectors might be an indecomposable vector), and also, as is easily verified, the set of decomposable vectors is not contained in any subspace of the space  $\Lambda^m(L)$  other than  $\Lambda^m(L)$  itself. In many problems, it is more natural to deal with vector spaces, and this is the reason for introducing the notion of a space  $\Lambda^m(L)$  that contains all  $m$ -vectors, including those that are indecomposable.

Let us note that the basis vectors  $\boldsymbol{\varphi}_I$  themselves are decomposable: they are determined by the conditions (10.20), which, as is easily verified, taking into account equality  $(F_J, \boldsymbol{\varphi}_I) = F_J(\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_m})$ , means that for a vector  $\mathbf{x} = \boldsymbol{\varphi}_I$ , we have the representation (10.23) for  $\mathbf{a}_1 = \mathbf{e}_{i_1}, \dots, \mathbf{a}_m = \mathbf{e}_{i_m}$ , that is,

$$\boldsymbol{\varphi}_I = \mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} \wedge \dots \wedge \mathbf{e}_{i_m}, \quad I = (i_1, \dots, i_m).$$

If  $\mathbf{e}_1, \dots, \mathbf{e}_n$  is a basis of the space  $L$ , then the vectors  $\mathbf{e}_{i_1} \wedge \dots \wedge \mathbf{e}_{i_m}$  for all possible increasing collections of indices  $(i_1, \dots, i_m)$  form a basis of the subspace  $\Lambda^m(L)$ , dual to the basis  $F_I$  of the space  $\Omega^m(L)$  that we considered above. Thus every  $m$ -vector is a linear combination of decomposable vectors.

The exterior product  $\mathbf{a}_1 \wedge \dots \wedge \mathbf{a}_m$  is a function of  $m$  vectors  $\mathbf{a}_i \in L$  with values in the space  $\Lambda^m(L)$ . Let us now establish some of its properties. The first two of these are an analogue of multilinearity, and the third is an analogue of antisymmetry, but taking into account that the exterior product is not a number, but a vector of the space  $\Lambda^m(L)$ .

*Property 10.11* For every  $i \in \{1, \dots, m\}$  and all vectors  $\mathbf{a}_i, \mathbf{b}, \mathbf{c} \in L$  the following relationship is satisfied:

$$\begin{aligned} & \mathbf{a}_1 \wedge \dots \wedge \mathbf{a}_{i-1} \wedge (\mathbf{b} + \mathbf{c}) \wedge \mathbf{a}_{i+1} \wedge \dots \wedge \mathbf{a}_m \\ &= \mathbf{a}_1 \wedge \dots \wedge \mathbf{a}_{i-1} \wedge \mathbf{b} \wedge \mathbf{a}_{i+1} \wedge \dots \wedge \mathbf{a}_m \\ &+ \mathbf{a}_1 \wedge \dots \wedge \mathbf{a}_{i-1} \wedge \mathbf{c} \wedge \mathbf{a}_{i+1} \wedge \dots \wedge \mathbf{a}_m. \end{aligned} \quad (10.24)$$

Indeed, by definition, the exterior product

$$\mathbf{a}_1 \wedge \dots \wedge \mathbf{a}_{i-1} \wedge (\mathbf{b} + \mathbf{c}) \wedge \mathbf{a}_{i+1} \wedge \dots \wedge \mathbf{a}_m$$

is a linear function on the space  $\Omega^m(L)$  associating with each function  $F \in \Omega^m(L)$ , the number  $F(\mathbf{a}_1, \dots, \mathbf{a}_{i-1}, \mathbf{b} + \mathbf{c}, \mathbf{a}_{i+1}, \dots, \mathbf{a}_m)$ . Since the function  $F$  is multilin-



ear, it follows that

$$\begin{aligned} F(\mathbf{a}_1, \dots, \mathbf{a}_{i-1}, \mathbf{b} + \mathbf{c}, \mathbf{a}_{i+1}, \dots, \mathbf{a}_m) \\ = F(\mathbf{a}_1, \dots, \mathbf{a}_{i-1}, \mathbf{b}, \mathbf{a}_{i+1}, \dots, \mathbf{a}_m) + F(\mathbf{a}_1, \dots, \mathbf{a}_{i-1}, \mathbf{c}, \mathbf{a}_{i+1}, \dots, \mathbf{a}_m), \end{aligned}$$

which proves equality (10.24).

The following two properties are just as easily verified.

*Property 10.12* For every number  $\alpha$  and all vectors  $\mathbf{a}_i \in \mathbb{L}$ , the following relationship holds:

$$\begin{aligned} \mathbf{a}_1 \wedge \dots \wedge \mathbf{a}_{i-1} \wedge (\alpha \mathbf{a}_i) \wedge \mathbf{a}_{i+1} \wedge \dots \wedge \mathbf{a}_m \\ = \alpha (\mathbf{a}_1 \wedge \dots \wedge \mathbf{a}_{i-1} \wedge \mathbf{a}_i \wedge \mathbf{a}_{i+1} \wedge \dots \wedge \mathbf{a}_m). \end{aligned} \quad (10.25)$$

*Property 10.13* For all pairs of indices  $r, s \in \{1, \dots, m\}$  and all vectors  $\mathbf{a}_i \in \mathbb{L}$ , the following relationship holds:

$$\begin{aligned} \mathbf{a}_1 \wedge \dots \wedge \mathbf{a}_{s-1} \wedge \mathbf{a}_s \wedge \mathbf{a}_{s+1} \wedge \dots \wedge \mathbf{a}_{r-1} \wedge \mathbf{a}_r \wedge \mathbf{a}_{r+1} \wedge \dots \wedge \mathbf{a}_m \\ = -\mathbf{a}_1 \wedge \dots \wedge \mathbf{a}_{s-1} \wedge \mathbf{a}_r \wedge \mathbf{a}_{s+1} \wedge \dots \\ \wedge \mathbf{a}_{r-1} \wedge \mathbf{a}_s \wedge \mathbf{a}_{r+1} \wedge \dots \wedge \mathbf{a}_m, \end{aligned} \quad (10.26)$$

that is, if any two vectors from among  $\mathbf{a}_1, \dots, \mathbf{a}_m$  change places, the exterior product changes sign.

If (as we assume) the numbers are elements of a field of characteristic different from 2 (for example,  $\mathbb{R}$  or  $\mathbb{C}$ ), then Property 10.13 yields the following corollary.

**Corollary 10.14** *If any two of the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m$  are equal, then  $\mathbf{a}_1 \wedge \dots \wedge \mathbf{a}_m = \mathbf{0}$ .*

Generalizing the definition given above, we may express Properties 10.11, 10.12, and 10.13 by saying that the exterior product  $\mathbf{a}_1 \wedge \dots \wedge \mathbf{a}_m$  is a multilinear antisymmetric function of the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m \in \mathbb{L}$  taking values in the space  $\Lambda^m(\mathbb{L})$ .

*Property 10.15* Vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m$  are linearly dependent if and only if

$$\mathbf{a}_1 \wedge \dots \wedge \mathbf{a}_m = \mathbf{0}. \quad (10.27)$$

*Proof* Let us assume that the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m$  are linearly dependent. Then one of them is a linear combination of the rest. Let it be the vector  $\mathbf{a}_m$  (the other cases are reduced to this one by a change in numeration). Then

$$\mathbf{a}_m = \alpha_1 \mathbf{a}_1 + \dots + \alpha_{m-1} \mathbf{a}_{m-1},$$

and on the basis of Properties 10.11 and 10.12, we obtain that

$$\begin{aligned} & \mathbf{a}_1 \wedge \cdots \wedge \mathbf{a}_{m-1} \wedge \mathbf{a}_m \\ &= \alpha_1(\mathbf{a}_1 \wedge \cdots \wedge \mathbf{a}_{m-1} \wedge \mathbf{a}_1) + \cdots + \alpha_{m-1}(\mathbf{a}_1 \wedge \cdots \wedge \mathbf{a}_{m-1} \wedge \mathbf{a}_{m-1}). \end{aligned}$$

In view of Corollary 10.14, each term on the right-hand side of this equality is equal to zero, and consequently, we have  $\mathbf{a}_1 \wedge \cdots \wedge \mathbf{a}_m = \mathbf{0}$ .

Let us assume now that the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m$  are linearly independent. We must prove that  $\mathbf{a}_1 \wedge \cdots \wedge \mathbf{a}_m \neq \mathbf{0}$ . Equality (10.27) would mean that the function  $\mathbf{a}_1 \wedge \cdots \wedge \mathbf{a}_m$  (as an element of the space  $\Lambda^m(L)$ ) assigns to an arbitrary function  $F \in \Omega^m(L)$ , the value  $F(\mathbf{a}_1, \dots, \mathbf{a}_m) = 0$ . However, in contradiction to this, it is possible to produce a function  $F \in \Omega^m(L)$  for which  $F(\mathbf{a}_1, \dots, \mathbf{a}_m) \neq 0$ . Indeed, let us represent the space  $L$  as a direct sum

$$L = \langle \mathbf{a}_1, \dots, \mathbf{a}_m \rangle \oplus L',$$

where  $L' \subset L$  is some subspace of dimension  $n - m$ , and for every vector  $\mathbf{z} \in L$ , let us consider the corresponding decomposition  $\mathbf{z} = \mathbf{x} + \mathbf{y}$ , where  $\mathbf{x} \in \langle \mathbf{a}_1, \dots, \mathbf{a}_m \rangle$  and  $\mathbf{y} \in L'$ . Finally, for vectors

$$\mathbf{z}_i = \alpha_{i1}\mathbf{a}_1 + \cdots + \alpha_{im}\mathbf{a}_m + \mathbf{y}_i, \quad \mathbf{y}_i \in L', i = 1, \dots, m,$$

let us define a function  $F$  by the condition  $F(\mathbf{z}_1, \dots, \mathbf{z}_m) = |\alpha_{ij}|$ . As we saw in Sect. 2.6, the determinant is a multilinear antisymmetric function of its rows. Moreover,  $F(\mathbf{a}_1, \dots, \mathbf{a}_m) = |E| = 1$ , which proves our assertion.  $\square$

Let  $L$  and  $M$  be arbitrary vector spaces, and let  $\mathcal{A} : L \rightarrow M$  be a linear transformation. It defines the transformation

$$\Omega^p(\mathcal{A}) : \Omega^p(M) \rightarrow \Omega^p(L), \quad (10.28)$$

which assigns to each antisymmetric function  $F(\mathbf{y}_1, \dots, \mathbf{y}_p)$  in the space  $\Omega^p(M)$ , an antisymmetric function  $G(\mathbf{x}_1, \dots, \mathbf{x}_p)$  in the space  $\Omega^p(L)$  by the formula

$$G(\mathbf{x}_1, \dots, \mathbf{x}_p) = F(\mathcal{A}(\mathbf{x}_1), \dots, \mathcal{A}(\mathbf{x}_p)), \quad \mathbf{x}_1, \dots, \mathbf{x}_p \in L. \quad (10.29)$$

A simple verification shows that this transformation is linear. Let us note that we have already met with such a transformation in the case  $m = 1$ , namely the dual transformation  $\mathcal{A}^* : M^* \rightarrow L^*$  (see Sect. 3.7). In the general case, passing to the dual spaces  $\Lambda^p(L) = \Omega^p(L)^*$  and  $\Lambda^p(M) = \Omega^p(M)^*$ , we define the linear transformation

$$\Lambda^p(\mathcal{A}) : \Lambda^p(L) \rightarrow \Lambda^p(M), \quad (10.30)$$

dual to the transformation (10.28).

Let us note the most important properties of the transformation (10.30).

**Lemma 10.16** *Let  $\mathcal{A} : L \rightarrow M$  and  $\mathcal{B} : M \rightarrow N$  be linear transformations of arbitrary vector spaces  $L, M, N$ . Then*

$$\Lambda^p(\mathcal{B}\mathcal{A}) = \Lambda^p(\mathcal{B})\Lambda^p(\mathcal{A}).$$

*Proof* In view of the definition (10.30) and the properties of dual transformations (formula (3.61)) established in Sect. 3.7, it suffices to ascertain that

$$\Omega^p(\mathcal{B}\mathcal{A}) = \Omega^p(\mathcal{A})\Omega^p(\mathcal{B}). \quad (10.31)$$

But equality (10.31) follows directly from the definition. Indeed, the transformation  $\Omega^p(\mathcal{A})$  maps the function  $F(y_1, \dots, y_p)$  in the space  $\Omega^p(M)$  to the function  $G(x_1, \dots, x_p)$  in  $\Omega^p(L)$  by formula (10.29). In just the same way, the transformation  $\Omega^p(\mathcal{B})$  maps the function  $H(z_1, \dots, z_p)$  in  $\Omega^p(N)$  to the function  $F(y_1, \dots, y_p)$  in  $\Omega^p(M)$  by the analogous formula

$$F(y_1, \dots, y_p) = H(\mathcal{B}(y_1), \dots, \mathcal{B}(y_p)), \quad y_1, \dots, y_p \in M. \quad (10.32)$$

Finally, the transformation  $\mathcal{B}\mathcal{A} : L \rightarrow N$  takes the function  $H(z_1, \dots, z_p)$  in the space  $\Omega^p(N)$  to the function  $G(x_1, \dots, x_p)$  in the space  $\Omega^p(L)$  by the formula

$$G(x_1, \dots, x_p) = H(\mathcal{B}\mathcal{A}(x_1), \dots, \mathcal{B}\mathcal{A}(x_p)), \quad x_1, \dots, x_p \in L. \quad (10.33)$$

Substituting into (10.33) the vector  $y_i = \mathcal{A}(x_i)$  and comparing the relationship thus obtained with (10.32), we obtain the required equality (10.31).  $\square$

**Lemma 10.17** *For all vectors  $x_1, \dots, x_p \in L$ , we have the equality*

$$\Lambda^p(\mathcal{A})(x_1 \wedge \dots \wedge x_p) = \mathcal{A}(x_1) \wedge \dots \wedge \mathcal{A}(x_p). \quad (10.34)$$

*Proof* Both sides of equality (10.34) are elements of the space  $\Lambda^p(M) = \Omega^p(M)^*$ , that is, they are linear functions on  $\Omega^p(M)$ . It suffices to verify that their application to any function  $F(y_1, \dots, y_p)$  in the space  $\Omega^p(M)$  gives one and the same result. But as follows from the definition, in both cases, this result is equal to  $F(\mathcal{A}(x_1), \dots, \mathcal{A}(x_p))$ .  $\square$

Finally, we shall prove a property of the exterior product that is sometimes called *universality*.

**Property 10.18** Any mapping that carries a vector  $[a_1, \dots, a_m]$  of some space  $M$  satisfying Properties 10.11, 10.12, 10.13 (p. 362) to  $m$  vectors  $a_1, \dots, a_m$  of the space  $L$  can be obtained from the exterior product  $a_1 \wedge \dots \wedge a_m$  by applying some uniquely defined linear transformation  $\mathcal{A} : \Lambda^m(L) \rightarrow M$ .

In other words, there exists a linear transformation  $\mathcal{A} : \Lambda^m(L) \rightarrow M$  such that for every collection  $a_1, \dots, a_m$  of vectors of the space  $L$ , we have the equality

$$[a_1, \dots, a_m] = \mathcal{A}(a_1 \wedge \dots \wedge a_m), \quad (10.35)$$

which can be represented by the following diagram:

$$\begin{array}{ccc}
 L^m & & \\
 \Lambda^m \downarrow & \searrow [\dots] & \\
 & M & \\
 & \nearrow \mathcal{A} & \\
 \Lambda^m(L) & & 
 \end{array} \tag{10.36}$$

In this diagram,  $[a_1, \dots, a_m] = \mathcal{A}(a_1 \wedge \dots \wedge a_m)$ .

Let us note that although  $L^m = L \times \dots \times L$  ( $m$ -fold product) is clearly a vector space, we by no means assert that the mapping

$$a_1, \dots, a_m \mapsto [a_1, \dots, a_m]$$

discussed in Property 10.18 is a linear transformation  $L^m \rightarrow M$ . In general, such is not the case. For example, the exterior product  $a_1 \wedge \dots \wedge a_m : L^m \rightarrow \Lambda^m(L)$  itself is not a linear transformation in the case that  $\dim L > m + 1$  and  $m > 1$ . Indeed, the image of the exterior product is the set of decomposable vectors described by their Plücker relations, which is not a vector subspace of  $\Lambda^m(L)$ .

*Proof of Property 10.18* We can construct a linear transformation  $\Psi : M^* \rightarrow \Omega^m(L)$  such that it maps every linear function  $f \in M^*$  to the function  $\Psi(f) \in \Omega^m(L)$  defined by the relationship

$$\Psi(f) = f([a_1, \dots, a_m]). \tag{10.37}$$

By Properties 10.11–10.13, which, by assumption, are satisfied by  $[a_1, \dots, a_m]$ , the mapping  $\Psi(f)$  thus constructed is a multilinear and antisymmetric function of  $a_1, \dots, a_m$ . Therefore,  $\Psi : M^* \rightarrow \Omega^m(L)$  is a linear transformation. Let us define  $\mathcal{A}$  as the dual mapping

$$\mathcal{A} = \Psi^* : \Lambda^m(L) = \Omega^m(L)^* \longrightarrow M = M^{**}.$$

By definition of the dual transformation (formula (3.58)), for every linear function  $F$  on the space  $\Omega^m(L)$ , its image  $\mathcal{A}(F)$  is a linear function on the space  $M^*$  such that  $\mathcal{A}(F)(f) = F(\Psi(f))$  for all  $f \in M^*$ . Applying formula (10.37) to the right-hand side of the last equality, we obtain the equality

$$\mathcal{A}(F)(f) = F(\Psi(f)) = F(f([a_1, \dots, a_m])). \tag{10.38}$$

Setting in (10.38) the function  $F(\Psi) = \Psi(a_1 \wedge \dots \wedge a_m)$ , that is,  $F = a_1 \wedge \dots \wedge a_m$ , we arrive at the relationship

$$\mathcal{A}(a_1 \wedge \dots \wedge a_m)(f) = f([a_1, \dots, a_m]), \tag{10.39}$$

whose left-hand side is an element of the space  $M^{**}$ , which is isomorphic to  $M$ .

Let us recall that the identification (isomorphism) of the spaces  $M^{**}$  and  $M$  can be obtained by mapping each vector  $\psi(f) \in M^{**}$  to the vector  $x \in M$  for which the equality  $f(x) = \psi(f)$  is satisfied for every linear function  $f \in M^*$ . Then formula (10.39) gives the relationship

$$f(\mathcal{A}(a_1 \wedge \cdots \wedge a_m)) = f([a_1, \dots, a_m]),$$

which is valid for every function  $f \in M^*$ . Consequently, from this we obtain the required relationship

$$\mathcal{A}(a_1 \wedge \cdots \wedge a_m) = [a_1, \dots, a_m]. \quad (10.40)$$

Equality (10.40) defines a linear transformation  $\mathcal{A}$  for all decomposable vectors  $x \in \Lambda^m(L)$ . But above, we saw that every  $m$ -vector is a linear combination of decomposable vectors. The transformation  $\mathcal{A}$  is linear, and therefore, it is uniquely defined for all  $m$ -vectors. Thus we obtain the required linear transformation  $\mathcal{A} : \Lambda^m(L) \rightarrow M$ .  $\square$

## 10.4 Exterior Algebras\*

In many branches of mathematics, an important role is played by the expression

$$a_1 \wedge \cdots \wedge a_m,$$

understood not so much as a function of  $m$  vectors  $a_1, \dots, a_m$  of the space  $L$  with values in  $\Lambda^m(L)$ , but more as the result of repeated ( $m$ -fold) application of the operation consisting in mapping two vectors  $x \in \Lambda^p(L)$  and  $y \in \Lambda^q(L)$  to the vector  $x \wedge y \in \Lambda^{p+q}(L)$ . For example, the expression  $a \wedge b \wedge c$  can then be calculated “by parts.” That is, it can be represented in the form  $a \wedge b \wedge c = (a \wedge b) \wedge c$  and computed by first calculating  $a \wedge b$ , and then  $(a \wedge b) \wedge c$ .

To accomplish this, we have first to define the function mapping two vectors  $x \in \Lambda^p(L)$  and  $y \in \Lambda^q(L)$  to the vector  $x \wedge y \in \Lambda^{p+q}(L)$ . As a first step, such a function  $x \wedge y$  will be defined for the case that the vector  $y \in \Lambda^q(L)$  is *decomposable*, that is, representable in the form

$$y = a_1 \wedge a_2 \wedge \cdots \wedge a_q, \quad a_i \in L. \quad (10.41)$$

Let us consider the mapping that assigns to  $p$  vectors  $b_1, \dots, b_p$  of the space  $L$  the vector

$$[b_1, \dots, b_p] = b_1 \wedge \cdots \wedge b_p \wedge a_1 \wedge \cdots \wedge a_q,$$

and let us apply to it Property 10.18 (universality) from the previous section. We thereby obtain the diagram

$$\begin{array}{ccc}
 L^p & & \\
 \Lambda^p \downarrow & \searrow [b_1, \dots, b_p] & \\
 & & \Lambda^{p+q}(L) \\
 & \nearrow \mathcal{A} & \\
 \Lambda^p(L) & & 
 \end{array} \tag{10.42}$$

In this diagram,

$$\mathcal{A}(b_1 \wedge \cdots \wedge b_p) = [b_1, \dots, b_p].$$

**Definition 10.19** Let  $y$  be a decomposable vector, that is, it can be written in the form (10.41). Then for every vector  $x \in \Lambda^p(L)$ , its image  $\mathcal{A}(x)$  for the transformation  $\mathcal{A} : \Lambda^p(L) \rightarrow \Lambda^{p+q}(L)$  constructed above is denoted by  $x \wedge y = x \wedge (a_1 \wedge \cdots \wedge a_q)$  and is called the *exterior product* of vectors  $x$  and  $y$ .

Thus as a first step, we defined  $x \wedge y$  in the case that the vector  $y$  is decomposable. In order to define  $x \wedge y$  for an arbitrary vector  $y \in \Lambda^q(L)$ , it suffices simply to repeat the same argument. Indeed, let us consider the mapping  $[a_1, \dots, a_q] : \Lambda^q(L) \rightarrow \Lambda^{p+q}(L)$  defined by the formula

$$[a_1, \dots, a_q] = x \wedge (a_1 \wedge \cdots \wedge a_q).$$

We again obtain, on the basis of Property 10.18, the same diagram:

$$\begin{array}{ccc}
 L^q & & \\
 \Lambda^q \downarrow & \searrow [a_1, \dots, a_q] & \\
 & & \Lambda^{p+q}(L) \\
 & \nearrow \mathcal{A} & \\
 \Lambda^q(L) & & 
 \end{array} \tag{10.43}$$

where the transformation  $\mathcal{A} : \Lambda^q(L) \rightarrow \Lambda^{p+q}(L)$  is defined by the formula

$$\mathcal{A}(a_1 \wedge \cdots \wedge a_q) = [a_1, \dots, a_q].$$

**Definition 10.20** For any vectors  $x \in \Lambda^p(L)$  and  $y \in \Lambda^q(L)$ , the *exterior product*  $x \wedge y$  is the vector  $\mathcal{A}(y) \in \Lambda^{p+q}(L)$  in diagram (10.43) constructed above.

Let us note some properties of the exterior product that follow from this definition.

*Property 10.21* For any vectors  $\mathbf{x}_1, \mathbf{x}_2 \in \Lambda^p(\mathbf{L})$  and  $\mathbf{y} \in \Lambda^q(\mathbf{L})$ , we have the relationship

$$(\mathbf{x}_1 + \mathbf{x}_2) \wedge \mathbf{y} = \mathbf{x}_1 \wedge \mathbf{y} + \mathbf{x}_2 \wedge \mathbf{y}.$$

Similarly, for any vectors  $\mathbf{x} \in \Lambda^p(\mathbf{L})$  and  $\mathbf{y} \in \Lambda^q(\mathbf{L})$  and any scalar  $\alpha$ , we have the relationship

$$(\alpha \mathbf{x}) \wedge \mathbf{y} = \alpha(\mathbf{x} \wedge \mathbf{y}).$$

Both equalities follow immediately from the definitions and the linearity of the transformation  $\mathcal{A}$  in diagram (10.43).

*Property 10.22* For any vectors  $\mathbf{x} \in \Lambda^p(\mathbf{L})$  and  $\mathbf{y}_1, \mathbf{y}_2 \in \Lambda^q(\mathbf{L})$ , we have the relationship

$$\mathbf{x} \wedge (\mathbf{y}_1 + \mathbf{y}_2) = \mathbf{x} \wedge \mathbf{y}_1 + \mathbf{x} \wedge \mathbf{y}_2.$$

Similarly, for any vectors  $\mathbf{x} \in \Lambda^p(\mathbf{L})$  and  $\mathbf{y} \in \Lambda^q(\mathbf{L})$  and any scalar  $\alpha$ , we have the relationship

$$\mathbf{x} \wedge (\alpha \mathbf{y}) = \alpha(\mathbf{x} \wedge \mathbf{y}).$$

Both equalities follow immediately from the definitions and the linearity of the transformations  $\mathcal{A}$  in diagrams (10.42) and (10.43).

*Property 10.23* For decomposable vectors  $\mathbf{x} = \mathbf{a}_1 \wedge \cdots \wedge \mathbf{a}_p$  and  $\mathbf{y} = \mathbf{b}_1 \wedge \cdots \wedge \mathbf{b}_q$ , we have the relationship

$$\mathbf{x} \wedge \mathbf{y} = \mathbf{a}_1 \wedge \cdots \wedge \mathbf{a}_p \wedge \mathbf{b}_1 \wedge \cdots \wedge \mathbf{b}_q.$$

This follows at once from the definition.

Let us note that we have actually defined the exterior product in such a way that Properties 10.21–10.23 are satisfied. Indeed, Property 10.23 defines the exterior product of decomposable vectors. And since every vector is a linear combination of decomposable vectors, it follows that Properties 10.21 and 10.22 define it in the general case. The property of universality of the exterior product has been necessary for verifying that the result  $\mathbf{x} \wedge \mathbf{y}$  does not depend on the choice of linear combinations of decomposable vectors that we use to represent the vectors  $\mathbf{x}$  and  $\mathbf{y}$ .

Finally, let us make note of the following equally simple property.

*Property 10.24* For any vectors  $\mathbf{x} \in \Lambda^p(\mathbf{L})$  and  $\mathbf{y} \in \Lambda^q(\mathbf{L})$ , we have the relationship

$$\mathbf{x} \wedge \mathbf{y} = (-1)^{pq} \mathbf{y} \wedge \mathbf{x}. \quad (10.44)$$

Both vectors on the right- and left-hand sides of equality (10.44) belong to the space  $\Lambda^{p+q}(\mathbf{L})$ , that is, by definition, they are linear functions on  $\Omega^{p+q}(\mathbf{L})$ . Since every vector is a linear combination of decomposable vectors, it suffices that we verify equality (10.44) for decomposable vectors.

Let  $\mathbf{x} = \mathbf{a}_1 \wedge \cdots \wedge \mathbf{a}_p$ ,  $\mathbf{y} = \mathbf{b}_1 \wedge \cdots \wedge \mathbf{b}_q$ , and let  $F$  be any vector of the space  $\Omega^{p+q}(\mathbf{L})$ , that is,  $F$  is an antisymmetric function of the vectors  $\mathbf{x}_1, \dots, \mathbf{x}_{p+q}$  in  $\mathbf{L}$ . Then equality (10.44) means that

$$F(\mathbf{a}_1, \dots, \mathbf{a}_p, \mathbf{b}_1, \dots, \mathbf{b}_q) = (-1)^{pq} F(\mathbf{b}_1, \dots, \mathbf{b}_q, \mathbf{a}_1, \dots, \mathbf{a}_p). \quad (10.45)$$

But equality (10.45) is an obvious consequence of the antisymmetry of the function  $F$ . Indeed, in order to place the vector  $\mathbf{b}_1$  in the first position on the left-hand side of (10.45), we must change the position of  $\mathbf{b}_1$  with each vector  $\mathbf{a}_1, \dots, \mathbf{a}_p$  in turn. One such transposition reverses the sign, and altogether, the transpositions multiply  $F$  by  $(-1)^p$ . Similarly, in order to place the vector  $\mathbf{b}_2$  in the second position on the left-hand side of (10.45), we also must execute  $p$  transpositions, and the value of  $F$  is again multiplied by  $(-1)^p$ . And in order to place all vectors  $\mathbf{b}_1, \dots, \mathbf{b}_q$  at the beginning, it is necessary to multiply  $F$  by  $(-1)^p$  a total of  $q$  times, and this ends up as (10.45).

Our next step consists in uniting all the sets  $\Lambda^p(\mathbf{L})$  into a single set  $\Lambda(\mathbf{L})$  and defining the exterior product for its elements. Here we encounter a special case of a very important algebraic notion, that of an *algebra*.<sup>2</sup>

**Definition 10.25** An *algebra* (over some field  $\mathbb{K}$ , which we shall consider to consist of numbers) is a vector space  $\mathbf{A}$  on which, besides the operations of addition of vectors and multiplication of a vector by a scalar, is also defined the operation  $\mathbf{A} \times \mathbf{A} \rightarrow \mathbf{A}$ , called the *product*, assigning to every pair of elements  $\mathbf{a}, \mathbf{b} \in \mathbf{A}$  the element  $\mathbf{ab} \in \mathbf{A}$  and satisfying the following conditions:

- (1) the distributive property: for all  $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbf{A}$ , we have the relationship

$$(\mathbf{a} + \mathbf{b})\mathbf{c} = \mathbf{ac} + \mathbf{bc}, \quad \mathbf{c}(\mathbf{a} + \mathbf{b}) = \mathbf{ca} + \mathbf{cb}; \quad (10.46)$$

- (2) for all  $\mathbf{a}, \mathbf{b} \in \mathbf{A}$  and every scalar  $\alpha \in \mathbb{K}$ , we have the relationship

$$(\alpha\mathbf{a})\mathbf{b} = \mathbf{a}(\alpha\mathbf{b}) = \alpha(\mathbf{ab}); \quad (10.47)$$

- (3) there exists an element  $\mathbf{e} \in \mathbf{A}$ , called the *identity*, such that for every  $\mathbf{a} \in \mathbf{A}$ , we have  $\mathbf{ea} = \mathbf{a}$  and  $\mathbf{ae} = \mathbf{a}$ .

Let us note that there can be only one identity element in an algebra. Indeed, if there existed another identity element  $\mathbf{e}'$ , then by definition, we would have the equalities  $\mathbf{ee}' = \mathbf{e}'$  and  $\mathbf{ee}' = \mathbf{e}$ , from which it follows that  $\mathbf{e} = \mathbf{e}'$ .

<sup>2</sup>This is not a very felicitous term, since it coincides with the name of a branch of mathematics, the one we are currently studying. But the term has taken root, and we are stuck with it.



As in any vector space, in an algebra we have, for every  $\mathbf{a} \in \mathbf{A}$ , the equality  $0 \cdot \mathbf{a} = \mathbf{0}$  (here the 0 on the left denotes the scalar zero in the field  $\mathbb{K}$ , while the  $\mathbf{0}$  on the right denotes the null element of the vector space  $\mathbf{A}$  that is an algebra).

If an algebra  $\mathbf{A}$  is finite-dimensional as a vector space and  $\mathbf{e}_1, \dots, \mathbf{e}_n$  is a basis of  $\mathbf{A}$ , then the elements  $\mathbf{e}_1, \dots, \mathbf{e}_n$  are said to form a *basis* of the algebra  $\mathbf{A}$ , where the number  $n$  is called its *dimension* and is denoted by  $\dim \mathbf{A} = n$ . For an algebra  $\mathbf{A}$  of finite dimension  $n$ , the product of two of its basis elements can be represented in the form

$$\mathbf{e}_i \mathbf{e}_j = \sum_{k=1}^n \alpha_{ij}^k \mathbf{e}_k, \quad i, j = 1, \dots, n, \quad (10.48)$$

where  $\alpha_{ij}^k \in \mathbb{K}$  are certain scalars.

The totality of all scalars  $\alpha_{ij}^k$  for all  $i, j, k = 1, \dots, n$  is called the *multiplication table* of the algebra  $\mathbf{A}$ , and it uniquely determines the product for all the elements of the algebra. Indeed, if  $\mathbf{x} = \lambda_1 \mathbf{e}_1 + \dots + \lambda_n \mathbf{e}_n$  and  $\mathbf{y} = \mu_1 \mathbf{e}_1 + \dots + \mu_n \mathbf{e}_n$ , then repeatedly applying the rules (10.46) and (10.47) and taking into account (10.48), we obtain

$$\mathbf{x} \mathbf{y} = \sum_{i,j,k=1}^n \lambda_i \mu_j \alpha_{ij}^k \mathbf{e}_k, \quad (10.49)$$

that is, the product  $\mathbf{x} \mathbf{y}$  is uniquely determined by the coordinates of the vectors  $\mathbf{x}, \mathbf{y}$  and the multiplication table of the algebra  $\mathbf{A}$ . And conversely, it is obvious that for any given multiplication table, formula (10.49) defines in an  $n$ -dimensional vector space an operation of multiplication satisfying all the requirements entering into the definition of an algebra, except, perhaps, property 3, which requires further consideration; that is, it converts this vector space into an algebra of the same dimension  $n$ .

**Definition 10.26** An algebra  $\mathbf{A}$  is said to be *associative* if for every collection of three elements  $\mathbf{a}, \mathbf{b}$ , and  $\mathbf{c}$ , we have the relationship

$$(\mathbf{a}\mathbf{b})\mathbf{c} = \mathbf{a}(\mathbf{b}\mathbf{c}). \quad (10.50)$$

The associative property makes it possible to calculate the product of any number of elements  $\mathbf{a}_1, \dots, \mathbf{a}_m$  of an algebra  $\mathbf{A}$  without indicating the arrangement of parentheses among them; see the discussion on p. xv. Clearly, it suffices to verify the associative property of a finite-dimensional algebra for elements of some basis.

We have already encountered some examples of algebras.

*Example 10.27* The algebra of all square matrices of order  $n$ . It has the finite dimension  $n^2$ , and as we saw in Sect. 2.9, it is associative.

*Example 10.28* The algebra of all polynomials in  $n > 0$  variables with numeric coefficients. This algebra is also associative, but its dimension is infinite.

Now we shall define for a vector space  $L$  of finite dimension  $n$  its *exterior algebra*  $\Lambda(L)$ . This algebra has many different applications (some of them will be discussed in the following section); its introduction is one more reason why in Sect. 10.3, we did not limit our consideration to decomposable vectors only, which were sufficient for describing vector subspaces.

Let us define the exterior algebra  $\Lambda(L)$  as a direct sum of spaces  $\Lambda^p(L)$ ,  $p \geq 0$ , which consist of more than just the one null vector, where  $\Lambda^0(L)$  is by definition equal to  $\mathbb{K}$ . Since as a result of the antisymmetry of the exterior product we have  $\Lambda^p(L) = (\mathbf{0})$  for all  $p > n$ , we obtain the following definition of an exterior algebra:

$$\Lambda(L) = \Lambda^0(L) \oplus \Lambda^1(L) \oplus \cdots \oplus \Lambda^n(L). \quad (10.51)$$

Thus every element  $u$  of the constructed vector space  $\Lambda(L)$  can be represented in the form  $u = u_0 + u_1 + \cdots + u_n$ , where  $u_i \in \Lambda^i(L)$ .

Our present goal is the definition of the exterior product in  $\Lambda(L)$ , which we denote by  $u \wedge v$  for arbitrary vectors  $u, v \in \Lambda(L)$ . We shall define the exterior product  $u \wedge v$  of vectors

$$u = u_0 + u_1 + \cdots + u_n, \quad v = v_0 + v_1 + \cdots + v_n, \quad u_i, v_i \in \Lambda^i(L),$$

as the element

$$u \wedge v = \sum_{i,j=0}^n u_i \wedge v_j,$$

where we use the fact that the exterior product  $u_i \wedge v_j$  is already defined as an element of the space  $\Lambda^{i+j}(L)$ . Thus

$$u \wedge v = w_0 + w_1 + \cdots + w_n, \quad \text{where } w_k = \sum_{i+j=k} u_i \wedge v_j, \quad w_k \in \Lambda^k(L).$$

A simple verification shows that for the exterior product thus defined, all the conditions for the definition of an algebra are satisfied. This follows at once from the properties of the exterior product  $x \wedge y$  of vectors  $x \in \Lambda^i(L)$  and  $y \in \Lambda^j(L)$  proved earlier. By definition,  $\Lambda^0(L) = \mathbb{K}$ , and the number 1 (the identity in the field  $\mathbb{K}$ ) is the identity in the exterior algebra  $\Lambda(L)$ .

**Definition 10.29** A finite-dimensional algebra  $A$  is called a *graded algebra* if there is given a decomposition of the vector space  $A$  into a direct sum of subspaces  $A_i \subset A$ ,

$$A = A_0 \oplus A_1 \oplus \cdots \oplus A_k, \quad (10.52)$$

and the following conditions are satisfied: for all vectors  $x \in A_i$  and  $y \in A_j$ , the product  $xy$  is in  $A_{i+j}$  if  $i + j \leq k$ , and  $xy = \mathbf{0}$  if  $i + j > k$ . Here the decomposition (10.52) is called a *grading*.

In this case,  $\dim A = \dim A_0 + \cdots + \dim A_k$ , and taking the union of the bases of the subspaces  $A_i$ , we obtain a basis of the space  $A$ . The decomposition (10.51) and the definition of the exterior product show that the exterior algebra  $\Lambda(L)$  is graded if the space  $L$  has finite dimension  $n$ . Since  $\Lambda^p(L) = (0)$  for all  $p > n$ , it follows that

$$\dim \Lambda(L) = \sum_{p=0}^n \dim \Lambda^p(L) = \sum_{p=0}^n C_n^p = 2^n.$$

In an arbitrary graded algebra  $A$  with grading (10.52), the elements of the subspace  $A_i$  are called *homogeneous elements of degree  $i$* , and for every  $u \in A_i$ , we write  $i = \deg u$ . One often encounters graded algebras of infinite dimension, and in this case, the grading (10.52) contains, in general, not a finite, but an infinite number of terms. For example, in the algebra of polynomials (Example 10.28), a grading is defined by the decomposition of a polynomial into homogeneous components.

Property (10.44) of the exterior product that we have proved shows that in an exterior algebra  $\Lambda(L)$ , we have for all homogeneous elements  $u$  and  $v$  the relationship

$$u \wedge v = (-1)^d v \wedge u, \quad \text{where } d = \deg u \deg v. \quad (10.53)$$

Let us prove that for every finite-dimensional vector space  $L$ , the exterior algebra  $\Lambda(L)$  is associative. As we noted above, it suffices to prove the associative property for some basis of the algebra. Such a basis can be constructed out of homogeneous elements, and we may even choose them to be decomposable. Thus we may suppose that the elements  $a, b, c \in \Lambda(L)$  are equal to

$$a = a_1 \wedge \cdots \wedge a_p, \quad b = b_1 \wedge \cdots \wedge b_q, \quad c = c_1 \wedge \cdots \wedge c_r,$$

and in this case, using the properties proved above, we obtain

$$a \wedge (b \wedge c) = a_1 \wedge \cdots \wedge a_p \wedge b_1 \wedge \cdots \wedge b_q \wedge c_1 \wedge \cdots \wedge c_r = (a \wedge b) \wedge c.$$

An associative graded algebra that satisfies relationship (10.53) for all pairs of homogeneous elements is called a *superalgebra*. Thus an exterior algebra  $\Lambda(L)$  of an arbitrary finite-dimensional vector space  $L$  is a superalgebra, and it is the most important example of this concept.

Let us now return to the exterior algebra  $\Lambda(L)$  of the finite-dimensional vector space  $L$ . Let us choose in it a convenient basis and determine its multiplication table.

Let us fix in the space  $L$  an arbitrary basis  $e_1, \dots, e_n$ . Since the elements  $\varphi_I = e_{i_1} \wedge \cdots \wedge e_{i_m}$  for all possible collections  $I = (i_1, \dots, i_m)$  in  $\vec{\mathbb{N}}_n^m$  form a basis of the space  $\Lambda^m(L)$ ,  $m > 0$ , it follows from decomposition (10.51) that a basis in  $\Lambda(L)$  is obtained as the union of the bases of the subspaces  $\Lambda^m(L)$  for all  $m = 1, \dots, n$  and the basis of the subspace  $\Lambda^0(L) = \mathbb{K}$ , consisting of a single nonnull scalar, for example 1. This means that all such elements  $\varphi_I$ ,  $I \in \vec{\mathbb{N}}_n^m$ ,  $m = 1, \dots, n$ , together with 1 form a basis of the exterior algebra  $\Lambda(L)$ . Since the

exterior product with 1 is trivial, it follows that in order to compose a multiplication table in the constructed basis, we must find the exterior product  $\varphi_I \wedge \varphi_J$  for all possible collections of indices  $I \in \overrightarrow{\mathbb{N}}_n^p$  and  $J \in \overrightarrow{\mathbb{N}}_n^q$  for all  $1 \leq p, q \leq n$ .

In view of Property 10.23 on page 369, the exterior product  $\varphi_I \wedge \varphi_J$  is equal to

$$\varphi_I \wedge \varphi_J = e_{i_1} \wedge \cdots \wedge e_{i_p} \wedge e_{j_1} \wedge \cdots \wedge e_{j_q}. \quad (10.54)$$

Here there are two possibilities. If the collections  $I$  and  $J$  contain at least one index in common, then by Corollary 10.14 (p. 363), the product (10.54) is equal to zero.

If, on the other hand,  $I \cap J = \emptyset$ , then we shall denote by  $K$  the collection in  $\overrightarrow{\mathbb{N}}_n^{p+q}$  comprising the indices belonging to the set  $I \cup J$ , that is, in other words,  $K$  is obtained by arranging the collection  $(i_1, \dots, i_p, j_1, \dots, j_q)$  in ascending order. Then, as is easily verified, the exterior product (10.54) differs from the element  $\varphi_K$ ,  $K \in \overrightarrow{\mathbb{N}}_n^{p+q}$ , belonging to the basis of the exterior algebra  $\Lambda(L)$  constructed above in that the indices of the collection  $I \cup J$  are not necessarily arranged in ascending order. In order to obtain from (10.54) the element  $\varphi_K$ ,  $K \in \overrightarrow{\mathbb{N}}_n^{p+q}$ , it is necessary to interchange the indices  $(i_1, \dots, i_p, j_1, \dots, j_q)$  in such a way that the resulting collection is increasing. Then by Theorems 2.23 and 2.25 from Sect. 2.6 and Property 10.13, according to which the exterior product changes sign under the transposition of any two vectors, we obtain that

$$\varphi_I \wedge \varphi_J = \varepsilon(I, J) \varphi_K, \quad K \in \overrightarrow{\mathbb{N}}_n^{p+q},$$

where the number  $\varepsilon(I, J)$  is equal to  $+1$  or  $-1$  depending on whether the number of transpositions necessary for passing from  $(i_1, \dots, i_p, j_1, \dots, j_q)$  to the collection  $K \in \overrightarrow{\mathbb{N}}_n^{p+q}$  is even or odd.

As a result, we see that in the constructed basis of the exterior algebra  $\Lambda(L)$ , the multiplication table assumes the following form:

$$\varphi_I \wedge \varphi_J = \begin{cases} 0, & \text{if } I \cap J \neq \emptyset, \\ \varepsilon(I, J) \varphi_K, & \text{if } I \cap J = \emptyset. \end{cases} \quad (10.55)$$

## 10.5 Appendix\*

The exterior product  $x \wedge y$  of vectors  $x \in \Lambda^p(L)$  and  $y \in \Lambda^q(L)$  defined in the previous section makes it possible in many cases to give simple proofs of assertions that we encountered earlier.

*Example 10.30* Let us consider the case  $p = n$ , using the notation and results of the previous section. As we have seen,  $\dim \Lambda^n(L) = C_n^n$ , and therefore, the space  $\Lambda^n(L)$  is one-dimensional, and each of its nonzero vectors constitutes a basis. If  $e$  is such a vector, then an arbitrary vector of the space  $\Lambda^n(L)$  can be written in the form  $\alpha e$

with a suitable scalar  $\alpha$ . Thus for any  $n$  vectors  $\mathbf{x}_1, \dots, \mathbf{x}_n$  of the space  $L$ , we obtain the relationship

$$\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_n = \alpha(\mathbf{x}_1, \dots, \mathbf{x}_n)\mathbf{e}, \quad (10.56)$$

where  $\alpha(\mathbf{x}_1, \dots, \mathbf{x}_n)$  is some function of  $n$  vectors taking numeric values from the field  $\mathbb{K}$ . By Properties 10.11, 10.12, and 10.13, this function is multilinear and antisymmetric.

Let us choose in the space  $L$  some basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  and set

$$\mathbf{x}_i = x_{i1}\mathbf{e}_1 + \cdots + x_{in}\mathbf{e}_n, \quad i = 1, \dots, n.$$

The choice of a basis defines an isomorphism of the space  $L$  and the space  $\mathbb{K}^n$  of rows of length  $n$ , in which the vector  $\mathbf{x}_i$  corresponds to the row  $(x_{i1}, \dots, x_{in})$ . Thus  $\alpha$  becomes a multilinear and antisymmetric function of  $n$  rows taking numeric values. By Theorem 2.15, the function  $\alpha(\mathbf{x}_1, \dots, \mathbf{x}_n)$  coincides up to a scalar multiple  $k(\mathbf{e})$  with the determinant of the square matrix of order  $n$  consisting of the coordinates  $x_{ij}$  of the vectors  $\mathbf{x}_1, \dots, \mathbf{x}_n$ :

$$\alpha(\mathbf{x}_1, \dots, \mathbf{x}_n) = k(\mathbf{e}) \cdot \begin{vmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nn} \end{vmatrix}. \quad (10.57)$$

The arbitrariness of the choice of coefficient  $k(\mathbf{e})$  in formula (10.57) corresponds to the arbitrariness of the choice of basis  $\mathbf{e}$  in the one-dimensional space  $\Lambda^n(L)$  (let us recall that the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the space  $L$  is fixed).

In particular, let us choose as basis of the space  $\Lambda^n(L)$  the vector

$$\mathbf{e} = \mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_n. \quad (10.58)$$

Vectors  $\mathbf{e}_1, \dots, \mathbf{e}_n$  are linearly independent. Therefore, by Property 10.15 (p. 363), the vector  $\mathbf{e}$  is nonnull. We therefore obviously obtain that  $k(\mathbf{e}) = 1$ . Indeed, since the coefficient  $k(\mathbf{e})$  in formula (10.57) is one and the same for all collections of vectors  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , we can calculate it by setting  $\mathbf{x}_i = \mathbf{e}_i$ ,  $i = 1, \dots, n$ . Comparing in this case formulas (10.56) and (10.58), we see that  $\alpha(\mathbf{e}_1, \dots, \mathbf{e}_n) = 1$ . Substituting this value into relationship (10.57) for  $\mathbf{x}_i = \mathbf{e}_i$ ,  $i = 1, \dots, n$ , and noting that the determinant on the right-hand side of (10.57) is the determinant of the identity matrix, that is, equal to 1, we conclude that  $k(\mathbf{e}) = 1$ .

Using definitions given earlier, we may associate the linear transformation  $\Lambda^n(\mathcal{A}) : \Lambda^n(L) \rightarrow \Lambda^n(L)$  with the linear transformation  $\mathcal{A} : L \rightarrow L$ . The transformation  $\mathcal{A}$  can be defined by indicating to which vectors  $\mathbf{x}_1, \dots, \mathbf{x}_n$  it takes the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the space  $L$ , that is, by specifying vectors  $\mathbf{x}_i = \mathcal{A}(\mathbf{e}_i)$ ,  $i = 1, \dots, n$ . By Lemma 10.17 (p. 365), we have the equality

$$\begin{aligned} \Lambda^n(\mathcal{A})(\mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_n) &= \mathcal{A}(\mathbf{e}_1) \wedge \cdots \wedge \mathcal{A}(\mathbf{e}_n) \\ &= \mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_n = \alpha(\mathbf{x}_1, \dots, \mathbf{x}_n)\mathbf{e}. \end{aligned} \quad (10.59)$$

On the other hand, as we know, all linear transformations of a one-dimensional space have the form  $x \mapsto \alpha x$ , where  $\alpha$  is some scalar equal to the determinant of the given transformation and independent of the choice of basis  $e$  in  $\Lambda^1(L)$ . Thus we obtain that  $(\Lambda^n(\mathcal{A}))(x) = \alpha x$ , where the scalar  $\alpha$  is equal to the determinant  $|\Lambda^n(\mathcal{A})|$  and clearly depends only on the transformation  $\mathcal{A}$  itself, that is, it is determined by the collection of vectors  $x_i = \mathcal{A}(e_i)$ ,  $i = 1, \dots, n$ . It is not difficult to see that this scalar  $\alpha$  coincides with the function  $\alpha(x_1, \dots, x_n)$  defined above. Indeed, let us choose in the space  $\Lambda^n(L)$  a basis  $e = e_1 \wedge \dots \wedge e_n$ . Then the required equality follows directly from formula (10.59).

Further, substituting into (10.59) expression (10.57) for  $\alpha(x_1, \dots, x_n)$ , taking into account that  $k(e) = 1$  and that the determinant on the right-hand side of (10.57) coincides with the determinant of the transformation  $\mathcal{A}$ , we obtain the following result:

$$\mathcal{A}(e_1) \wedge \dots \wedge \mathcal{A}(e_n) = |\mathcal{A}|(e_1 \wedge \dots \wedge e_n). \quad (10.60)$$

This relationship gives the most invariant definition of the determinant of a linear transformation among all those that we have encountered.

We obtained relationship (10.60) for an arbitrary basis  $e_1, \dots, e_n$  of the space  $L$ , that is, for any  $n$  linearly independent vectors of the space. But it is also true for any  $n$  linearly dependent vectors  $a_1, \dots, a_n$  of this space. Indeed, in this case, the vectors  $\mathcal{A}(a_1), \dots, \mathcal{A}(a_n)$  are clearly also linearly dependent, and by Property 10.15, both exterior products  $a_1 \wedge \dots \wedge a_n$  and  $\mathcal{A}(a_1) \wedge \dots \wedge \mathcal{A}(a_n)$  are equal to zero. Thus for any  $n$  vectors  $a_1, \dots, a_n$  of the space  $L$  and any linear transformation  $\mathcal{A} : L \rightarrow L$ , we have the relationship

$$\mathcal{A}(a_1) \wedge \dots \wedge \mathcal{A}(a_n) = |\mathcal{A}|(a_1 \wedge \dots \wedge a_n). \quad (10.61)$$

In particular, if  $\mathcal{B} : L \rightarrow L$  is some other linear transformation, then formula (10.60) for the transformation  $\mathcal{B}\mathcal{A} : L \rightarrow L$  gives the analogous equality

$$(\mathcal{B}\mathcal{A}(e_1) \wedge \dots \wedge \mathcal{B}\mathcal{A}(e_n)) = |\mathcal{B}\mathcal{A}|(e_1 \wedge \dots \wedge e_n).$$

On the other hand, from the same formula we obtain that

$$\begin{aligned} (\mathcal{B}(\mathcal{A}(e_1)) \wedge \dots \wedge \mathcal{B}(\mathcal{A}(e_n))) &= |\mathcal{B}|(\mathcal{A}(e_1) \wedge \dots \wedge \mathcal{A}(e_n)) \\ &= |\mathcal{B}||\mathcal{A}|(e_1 \wedge \dots \wedge e_n). \end{aligned}$$

Hence it follows that  $|\mathcal{B}\mathcal{A}| = |\mathcal{B}| \cdot |\mathcal{A}|$ . This is almost a “tautological” proof of Theorem 2.54 on the determinant of the product of square matrices.

The arguments that we have presented acquire a more concrete character if  $L$  is an oriented Euclidean space. Then as the basis  $e_1, \dots, e_n$  in  $L$  we may choose an orthonormal and positively oriented basis. In this case, the basis (10.58) in  $\Lambda^n(L)$  is uniquely defined, that is, it does not depend on the choice of basis  $e_1, \dots, e_n$ . Indeed, if  $e'_1, \dots, e'_n$  is another such basis in  $L$ , then as we know, there exists a linear transformation  $\mathcal{A} : L \rightarrow L$  such that  $e'_i = \mathcal{A}(e_i)$ ,  $i = 1, \dots, n$ , and furthermore, the transformation  $\mathcal{A}$  is orthogonal and proper. But then  $|\mathcal{A}| = 1$ , and formula (10.60) shows that  $e'_1 \wedge \dots \wedge e'_n = e_1 \wedge \dots \wedge e_n$ .

*Example 10.31* Let us show how from the given considerations, we obtain a proof of the *Cauchy–Binet formula*, which was stated but not proved in Sect. 2.9.

Let us recall that in that section, we considered the product of two matrices  $B$  and  $A$ , the first of type  $(m, n)$ , and the second of type  $(n, m)$ , so that  $BA$  is a square matrix of order  $m$ . We are required to obtain an expression for the determinant  $|BA|$  in terms of the *associated* minors of the matrices  $B$  and  $A$ . Minors of the matrices  $B$  and  $A$  are said to be associated if they are of the same order, namely the minimum of  $n$  and  $m$ , and are located in the columns (of matrix  $B$ ) and rows (of matrix  $A$ ) of identical indices. The Cauchy–Binet formula asserts that the determinant  $|BA|$  is equal to 0 if  $n < m$ , and that  $|BA|$  is equal to the sum of the pairwise products over all the associated minors of order  $m$  if  $n \geq m$ .

Since every matrix is the matrix of some linear transformation of vector spaces of suitable dimensions, we may formulate this problem as a question of the determinant of the product of linear transformations  $\mathcal{A} : M \rightarrow L$  and  $\mathcal{B} : L \rightarrow M$ , where  $\dim L = n$  and  $\dim M = m$ . Here it is assumed that we have chosen a basis  $e_1, \dots, e_m$  in the space  $M$  and a basis  $f_1, \dots, f_n$  in the space  $L$  such that the transformations  $\mathcal{A}$  and  $\mathcal{B}$  have matrices  $A$  and  $B$  respectively in these bases. Then  $\mathcal{B}\mathcal{A}$  will be a linear transformation of the space  $M$  into itself with determinant  $|\mathcal{B}\mathcal{A}| = |BA|$ .

Let us first prove that  $|BA| = 0$  if  $n < m$ . Since the image of the transformation,  $\mathcal{B}\mathcal{A}(M)$ , is a subset of  $\mathcal{B}(L)$  and  $\dim \mathcal{B}(L) \leq \dim L$ , it follows that in the case under consideration, we have the inequality

$$\dim(\mathcal{B}\mathcal{A}(M)) \leq \dim \mathcal{B}(L) \leq \dim L = n < m = \dim M,$$

from which it follows that the image of the transformation  $\mathcal{B}\mathcal{A} : M \rightarrow M$  is not equal to the entire space  $M$ , that is, the transformation  $\mathcal{B}\mathcal{A}$  is singular. This means that  $|\mathcal{B}\mathcal{A}| = 0$ , that is,  $|BA| = 0$ .

Now let us consider the case  $n \geq m$ . Using Lemmas 10.16 and 10.17 from Sect. 10.3 with  $p = m$ , we obtain for the vectors of the basis  $e_1, \dots, e_m$  of the space  $M$  the relationship

$$\begin{aligned} \Lambda^m(\mathcal{B}\mathcal{A})(e_1 \wedge \dots \wedge e_m) &= \Lambda^m(\mathcal{B})\Lambda^m(\mathcal{A})(e_1 \wedge \dots \wedge e_m) \\ &= \Lambda^m(\mathcal{B})(\mathcal{A}(e_1) \wedge \dots \wedge \mathcal{A}(e_m)). \end{aligned} \quad (10.62)$$

The vectors  $\mathcal{A}(e_1), \dots, \mathcal{A}(e_m)$  are contained in the space  $L$  of dimension  $n$ , and their coordinates in the basis  $f_1, \dots, f_n$ , being written in column form, form the matrix  $A$  of the transformation  $\mathcal{A} : M \rightarrow L$ . Let us now write the coordinates of the vectors  $\mathcal{A}(e_1), \dots, \mathcal{A}(e_m)$  in row form. We thereby obtain the transpose matrix  $A^*$  of type  $(m, n)$ . Applying formula (10.22) to the vectors  $\mathcal{A}(e_1), \dots, \mathcal{A}(e_m)$ , we obtain the equality

$$\mathcal{A}(e_1) \wedge \dots \wedge \mathcal{A}(e_m) = \sum_{I \subset \vec{\mathbb{N}}_n^m} M_I \varphi_I \quad (10.63)$$

with the functions  $\varphi_I$  defined by formula (10.20). In the expression (10.63), according to our definition,  $M_I$  is the minor of the matrix  $A^*$  occupying columns

$i_1, \dots, i_m$ . It is obvious that such a minor  $M_I$  of the matrix  $A^*$  coincides with the minor of the matrix  $A$  occupying rows with the same indices  $i_1, \dots, i_m$ . Thus we may assume that in the sum on the right-hand side of (10.63),  $M_I$  are the minors of order  $m$  of the matrix  $A$  corresponding to all possible ordered collections  $I = (i_1, \dots, i_m)$  of indices of its rows.

Relationships (10.62) and (10.63) together give the equality

$$\Lambda^m(\mathcal{BA})(e_1 \wedge \dots \wedge e_m) = \Lambda^m(\mathcal{B}) \left( \sum_{I \subset \vec{\mathbb{N}}_n^m} M_I \varphi_I \right). \quad (10.64)$$

Let us denote by  $M_I$  and  $N_I$  the associated minors of the matrices  $A$  and  $B$ . This means that the minor  $M_I$  occupies the rows of the matrix  $A$  with indices  $I = (i_1, \dots, i_m)$ , and the minor  $N_I$  occupies the columns of the matrix  $B$  with the same indices. Let us consider the restriction of the linear transformation  $\mathcal{B} : \mathbb{L} \rightarrow \mathbb{M}$  to the subspace  $\langle f_{i_1}, \dots, f_{i_m} \rangle$ . By the definition of the functions  $\varphi_I$ , we obtain that

$$\Lambda^m(\mathcal{B})(\varphi_I) = \mathcal{B}(f_{i_1}) \wedge \dots \wedge \mathcal{B}(f_{i_m}) = N_I(e_1 \wedge \dots \wedge e_m).$$

From this, taking into account formula (10.64), follows the relationship

$$\begin{aligned} \Lambda^m(\mathcal{BA})(e_1 \wedge \dots \wedge e_m) &= \Lambda^m(\mathcal{B}) \left( \sum_{I \subset \vec{\mathbb{N}}_n^m} M_I \varphi_I \right) \\ &= \sum_{I \subset \vec{\mathbb{N}}_n^m} M_I \Lambda^m(\mathcal{B})(\varphi_I) \\ &= \left( \sum_{I \subset \vec{\mathbb{N}}_n^m} M_I N_I \right) (e_1 \wedge \dots \wedge e_m). \end{aligned}$$

On the other hand, by Lemma 10.17 and formula (10.60), we have

$$\Lambda^m(\mathcal{BA})(e_1 \wedge \dots \wedge e_m) = \mathcal{BA}(e_1) \wedge \dots \wedge \mathcal{BA}(e_m) = |\mathcal{BA}|(e_1 \wedge \dots \wedge e_m).$$

The last two equalities give us the relationship

$$|\mathcal{BA}| = \sum_{I \subset \vec{\mathbb{N}}_n^m} M_I N_I,$$

which, taking into account the equality  $|\mathcal{BA}| = |BA|$ , coincides with the Cauchy–Binet formula for the case  $n \geq m$ .

*Example 10.32* Let us derive the formula for the determinant of a square matrix  $A$  that generalizes the well-known formula for the expansion of the determinant along the  $j$ th column:

$$|A| = a_{1j} A_{1j} + a_{2j} A_{2j} + \dots + a_{nj} A_{nj}, \quad (10.65)$$



where  $A_{ij}$  is the cofactor of the element  $a_{ij}$ , that is, the number  $(-1)^{i+j} M_{ij}$ , and  $M_{ij}$  is the minor obtained by deleting this element from the matrix  $A$  along with the entire row and column at whose intersection it is located. The generalization consists in the fact that now we shall write down an analogous expansion of the determinant not along a single column, but along several, thereby generalizing in a suitable way the notion of the cofactor.

Let us consider a certain collection  $I \in \vec{\mathbb{N}}_n^m$ , where  $m$  is a natural number in the range 1 to  $n - 1$ . Let us denote by  $\bar{I}$  the collection obtained from  $(1, \dots, n)$  by discarding all indices entering into  $I$ . Clearly,  $\bar{I} \in \vec{\mathbb{N}}_n^{n-m}$ . Let us denote by  $|I|$  the sum of all indices entering into the collection  $I$ , that is, we shall set  $|I| = i_1 + \dots + i_m$ .

Let  $A$  be an arbitrary square matrix of order  $n$ , and let  $I = (i_1, \dots, i_m)$  and  $J = (j_1, \dots, j_m)$  be two collections of indices in  $\vec{\mathbb{N}}_n^m$ . For the minor  $M_{IJ}$  occupying the rows with indices  $i_1, \dots, i_m$  and columns with indices  $j_1, \dots, j_m$ , let us call the number

$$A_{IJ} = (-1)^{|I|+|J|} M_{\bar{I}\bar{J}} \quad (10.66)$$

the *cofactor*. It is easy to see that the given definition is indeed a generalization of that given in Chap. 2 of the cofactor of a single element  $a_{ij}$  for which  $m = 1$  and the collections  $I = (i)$ ,  $J = (j)$  each consist of a single index.

**Theorem 10.33** (Laplace's theorem) *The determinant of a matrix  $A$  is equal to the sum of the products of all minors occupying any  $m$  given columns (or rows) by their cofactors:*

$$|A| = \sum_{J \in \vec{\mathbb{N}}_n^m} M_{IJ} A_{IJ} = \sum_{I \in \vec{\mathbb{N}}_n^m} M_{IJ} A_{IJ},$$

where the number  $m$  can be arbitrarily chosen in the range 1 to  $n - 1$ .

**Remark 10.34** For  $m = 1$  and  $m = n - 1$ , Laplace's theorem gives formula (10.65) for the expansion of the determinant along a column and the analogous formula for expansion along a row. However, only in the general case is it possible to focus our attention on the symmetry between the minors of order  $m$  and those of order  $n - m$ .

**Proof of Theorem 10.33** Let us first of all note that since for the transpose matrix, its rows are converted into columns while the determinant is unchanged, it suffices to provide a proof for only one of the given equalities. For definiteness, let us prove the first—the formula for the expansion of the determinant  $|A|$  along  $m$  columns.

Let us consider a vector space  $L$  of dimension  $n$  and an arbitrary basis  $e_1, \dots, e_n$  of  $L$ . Let  $\mathcal{A} : L \rightarrow L$  be a linear transformation having in this basis the matrix  $A$ . Let us apply to the vectors of this basis a permutation such that the first  $m$  positions are occupied by the vectors  $e_{i_1}, \dots, e_{i_m}$ , the remaining  $n - m$  positions by the vectors  $e_{i_{m+1}}, \dots, e_{i_n}$ . In the basis thus obtained, the determinant of the transformation  $\mathcal{A}$

will again be equal to  $|A|$ , since the determinant of the matrix of a transformation  $\mathcal{A}$  does not depend on the choice of basis. Using formula (10.60), we obtain

$$\begin{aligned} \mathcal{A}(e_{i_1}) \wedge \cdots \wedge \mathcal{A}(e_{i_m}) \wedge \mathcal{A}(e_{i_{m+1}}) \wedge \cdots \wedge \mathcal{A}(e_{i_n}) \\ = |A|(e_{i_1} \wedge \cdots \wedge e_{i_m} \wedge e_{i_{m+1}} \wedge \cdots \wedge e_{i_n}) = |A|(\varphi_I \wedge \varphi_{\bar{I}}). \end{aligned} \quad (10.67)$$

Let us calculate the left-hand side of relationship (10.67), applying formula (10.22) to the two different groups of vectors.

First, let us set  $a_1 = \mathcal{A}(e_{i_1}), \dots, a_m = \mathcal{A}(e_{i_m})$ . Then from (10.22), we obtain

$$\mathcal{A}(e_{i_1}) \wedge \cdots \wedge \mathcal{A}(e_{i_m}) = \sum_{J \in \vec{\mathbb{N}}_n^m} M_{IJ} \varphi_J, \quad (10.68)$$

where  $I = (i_1, \dots, i_m)$ , and  $J$  runs through all collections from the set  $\vec{\mathbb{N}}_n^m$ .

Now let replace the number  $m$  by  $n - m$  in (10.22) and apply the formula thus obtained to the vectors  $a_1 = \mathcal{A}(e_{i_{m+1}}), \dots, a_{n-m} = \mathcal{A}(e_{i_n})$ . As a result, we obtain the equality

$$\mathcal{A}(e_{i_{m+1}}) \wedge \cdots \wedge \mathcal{A}(e_{i_n}) = \sum_{J' \in \vec{\mathbb{N}}_n^{n-m}} M_{\bar{I}J'} \varphi_{J'}, \quad (10.69)$$

where  $\bar{I} = (i_{m+1}, \dots, i_n)$ , and  $J'$  runs through all collections in the set  $\vec{\mathbb{N}}_n^{n-m}$ .

Substituting the expressions (10.68) and (10.69) into the left-hand side of (10.67), we obtain the equality

$$\sum_{J \in \vec{\mathbb{N}}_n^m} \sum_{J' \in \vec{\mathbb{N}}_n^{n-m}} M_{IJ} M_{\bar{I}J'} \varphi_J \wedge \varphi_{J'} = |A|(\varphi_I \wedge \varphi_{\bar{I}}). \quad (10.70)$$

Let us calculate the exterior product  $\varphi_I \wedge \varphi_{\bar{I}}$  for  $p = m$  and  $q = n - m$ , making use of the multiplication table (10.55) that was obtained at the end of the previous section. In this case, it is obvious that the collection  $K$  obtained by the union of  $I$  and  $\bar{I}$  is equal to  $(1, \dots, n)$ , and we have only to calculate the number  $\varepsilon(I, \bar{I}) = \pm 1$ , which depends on whether the number of transpositions to get from  $(i_1, \dots, i_m, i_{m+1}, \dots, i_n)$  to  $K = (1, \dots, n)$  is even or odd. It is not difficult to see (using, for example, the same reasoning as in Sect. 2.6) that  $\varepsilon(I, \bar{I})$  is equal to the number of pairs  $(i, \bar{i})$ , where  $i \in I$  and  $\bar{i} \in \bar{I}$ , for which the indices  $i$  and  $\bar{i}$  are in reverse order (form an inversion), that is,  $i > \bar{i}$ . By definition, all indices less than  $i_1$  appear in  $\bar{I}$ , and consequently, they form an inversion with  $i_1$ . This gives us  $i_1 - 1$  pairs. Further, all numbers less than  $i_2$  and belonging to  $\bar{I}$  form an inversion with index  $i_2$ , that is, all numbers less than  $i_2$  with the exception of  $i_1$ , which belongs to  $I$  and not  $\bar{I}$ . This gives  $i_2 - 2$  pairs.

Continuing in this way to the end, we obtain that the number of pairs  $(i, \bar{i})$  forming an inversion is equal to  $(i_1 - 1) + (i_2 - 2) + \cdots + (i_m - m)$ , that is, equal to  $|I| - \mu$ , where  $\mu = 1 + \cdots + m = \frac{1}{2}m(m+1)$ . Consequently, we finally obtain the formula  $\varphi_I \wedge \varphi_{\bar{I}} = (-1)^{|I|-\mu} \varphi_K$ , where  $K = (1, \dots, n)$ .

The exterior product  $\varphi_J \wedge \varphi_{J'}$  is equal to zero for all  $J$  and  $J'$ , with the exception only of the case that  $J' = \overline{J}$ , that is, the collections  $J$  and  $J'$  are disjoint and complement each other. By what we have said above,  $\varphi_J \wedge \varphi_{\overline{J}} = (-1)^{|J|-\mu} \varphi_K$ . Thus from (10.70) we obtain the equality

$$\sum_{J \in \vec{\mathbb{N}}_n^m} M_{IJ} M_{\overline{I}J} (-1)^{|J|-\mu} \varphi_K = |A| (-1)^{|I|-\mu} \varphi_K. \quad (10.71)$$

Multiplying both sides of equality (10.71) by the number  $(-1)^{|I|+\mu}$ , taking into account the obvious identity  $(-1)^{2|I|} = 1$ , we finally obtain

$$\sum_{J \in \vec{\mathbb{N}}_n^m} M_{IJ} M_{\overline{I}J} (-1)^{|I|+|J|} = |A|,$$

which, taking into account definition (10.66), gives us the required equality.  $\square$

*Example 10.35* We began this section with Example 10.30, in which we investigated in detail the space  $\Lambda^p(L)$  for  $p = n$ . Let us now consider the case  $p = n - 1$ . As a result of the general relationship  $\dim \Lambda^p(L) = C_n^p$ , we obtain that  $\dim \Lambda^{n-1}(L) = n$ .

Having chosen an arbitrary basis  $e_1, \dots, e_n$  in the space  $L$ , we assign to every vector  $z \in \Lambda^{n-1}(L)$  the linear function  $f(x)$  on  $L$  defined by the condition

$$z \wedge x = f(x)(e_1 \wedge \dots \wedge e_n), \quad x \in L.$$

For this, it is necessary to recall that  $z \wedge x$  belongs to the one-dimensional space  $\Lambda^n(L)$ , and the vector  $e_1 \wedge \dots \wedge e_n$  constitutes there a basis. The linearity of the function  $f(x)$  follows from the properties of the exterior product proved above. Let us verify that the linear transformation

$$\mathcal{F} : \Lambda^{n-1}(L) \rightarrow L^*$$

thus constructed is an isomorphism. Since  $\dim \Lambda^{n-1}(L) = \dim L^* = n$ , to show this, it suffices to verify that the kernel of the transformation  $\mathcal{F}$  is equal to  $(0)$ . As we know, it is possible to select as the basis of the space  $\Lambda^{n-1}(L)$  the vectors

$$e_{i_1} \wedge e_{i_2} \wedge \dots \wedge e_{i_{n-1}}, \quad i_k \in \{1, \dots, n\},$$

uniquely up to a permutation of the collection  $(i_1, \dots, i_{n-1})$ ; these are all the numbers  $(1, \dots, n)$  except for one. This means that as the basis  $\Lambda^{n-1}(L)$  one can choose the vectors

$$u_i = e_1 \wedge \dots \wedge e_{i-1} \wedge \check{e}_i \wedge e_{i+1} \wedge \dots \wedge e_n, \quad i = 1, \dots, n. \quad (10.72)$$

It is clear that  $u_i \wedge e_j = 0$  if  $i \neq j$ , and  $u_i \wedge e_i = \pm e_1 \wedge \dots \wedge e_n$  for all  $i = 1, \dots, n$ .

Let us assume that  $z \in \Lambda^{n-1}(L)$  is a nonnull vector such that its associated linear function  $f(x)$  is equal to zero for every  $x \in L$ . Let us set  $z = z_1 u_1 + \dots + z_n u_n$ .

Then from our assumption, it follows that  $z \wedge x = \mathbf{0}$  for all  $x \in L$ , and in particular, for the vectors  $e_1, \dots, e_n$ . It is easy to see that from this follow the equalities  $z_1 = 0, \dots, z_n = 0$  and hence  $z = \mathbf{0}$ .

The constructed isomorphism  $\mathcal{F} : \Lambda^{n-1}(L) \rightarrow L^*$  is a refinement of the following fact that we encountered earlier: the Plücker coordinates of a hyperplane can be arbitrary numbers; in this dimension, the Plücker relations do not yet appear.

Let us now assume that the space  $L$  is an oriented Euclidean space. On the one hand, this determines a fixed basis (10.58) in  $\Lambda^n(L)$  if  $e_1, \dots, e_n$  is an arbitrary positively oriented orthonormal basis of  $L$ , so that the isomorphism  $\mathcal{F} : \Lambda^{n-1}(L) \rightarrow L^*$  constructed above is uniquely determined. On the other hand, for a Euclidean space, there is defined the standard isomorphism  $L^* \xrightarrow{\sim} L$ , which does not require the selection of any basis at all in  $L$  (see p. 214). Combining these two isomorphisms, we obtain the isomorphism

$$\mathcal{G} : \Lambda^{n-1}(L) \xrightarrow{\sim} L,$$

which assigns to the element  $z \in \Lambda^{n-1}(L)$  the vector  $x \in L$  such that

$$z \wedge y = (x, y)(e_1 \wedge \dots \wedge e_n) \quad (10.73)$$

for every vector  $y \in L$  and for the positively oriented orthonormal basis  $e_1, \dots, e_n$ , where  $(x, y)$  denotes the inner product in the space  $L$ .

Let us consider this isomorphism in greater detail. We saw earlier that the vectors  $u_i$  determined by formula (10.72) form a basis of the space  $\Lambda^{n-1}(L)$ . To describe the constructed isomorphism, it suffices to determine which vector  $b \in L$  corresponds to the vector  $a_1 \wedge \dots \wedge a_{n-1}$ ,  $a_i \in L$ . We may suppose that the vectors  $a_1, \dots, a_{n-1}$  are linearly independent, since otherwise, the vector  $a_1 \wedge \dots \wedge a_{n-1}$  would equal  $\mathbf{0}$ , and therefore to it would correspond the vector  $b = \mathbf{0}$ . Taking into account formula (10.73), this correspondence implies the equality

$$(b, y)(e_1 \wedge \dots \wedge e_n) = a_1 \wedge \dots \wedge a_{n-1} \wedge y, \quad (10.74)$$

satisfied by all  $y \in L$ . Since the vector on the right-hand side of (10.74) is the null vector if  $y$  belongs to the subspace  $L_1 = \langle a_1, \dots, a_{n-1} \rangle$ , we may assume that  $b \in L_1^\perp$ .

Now we must recall that we have an orientation and consider  $L$  and  $L_1$  to be oriented (it is easy to ascertain that the orientation of the space  $L$  does not determine a natural orientation of the subspace  $L_1$ , and so we must choose and fix the orientation of  $L_1$  separately). Then we may choose the basis  $e_1, \dots, e_n$  in such a way that it is orthonormal and positively oriented and also such that the first  $n-1$  vectors  $e_1, \dots, e_{n-1}$  belong to the subspace  $L_1$ , and also define in it an orthonormal and positively oriented basis (it is always possible to attain this, possibly after replacing the vector  $e_n$  with its opposite).

Since the vector  $b$  is contained in the one-dimensional subspace  $L_1^\perp = \langle e_n \rangle$ , it follows that  $b = \beta e_n$ . Using the previous arguments, we obtain that

$$a_1 \wedge \dots \wedge a_{n-1} = v(a_1, \dots, a_{n-1})e_n,$$

where  $v(\mathbf{a}_1, \dots, \mathbf{a}_{n-1})$  is the oriented volume of the parallelepiped spanned by the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_{n-1}$  (see the definition on p. 221). This observation determines the number  $\beta$ .

Indeed, substituting the vector  $\mathbf{y} = \mathbf{e}_n$  into (10.74) and taking into account the fact that the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  was chosen to be orthonormal and positively oriented (from which follows, in particular, the equality  $v(\mathbf{e}_1 \wedge \dots \wedge \mathbf{e}_n) = 1$ ), we obtain the relationship

$$\beta v = v(\mathbf{a}_1, \dots, \mathbf{a}_{n-1}, \mathbf{e}_n) = v(\mathbf{a}_1, \dots, \mathbf{a}_{n-1}).$$

Thus the isomorphism  $\mathcal{G}$  constructed above assigns to the vector  $\mathbf{a}_1 \wedge \dots \wedge \mathbf{a}_{n-1}$  the vector  $\mathbf{b} = v(\mathbf{a}_1, \dots, \mathbf{a}_{n-1})\mathbf{e}_n$ , where  $\mathbf{e}_n$  is the unit vector on the line  $\mathbb{L}_1^\perp$ , chosen with the sign making the basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the space  $\mathbb{L}$  orthonormal and positively oriented. As is easily verified, this is equivalent to the requirement that the basis  $\mathbf{a}_1, \dots, \mathbf{a}_{n-1}, \mathbf{e}_n$  be positively oriented.

The final result is contained in the following theorem.

**Theorem 10.36** *For every oriented Euclidean space  $\mathbb{L}$ , the isomorphism*

$$\mathcal{G} : \Lambda^{n-1}(\mathbb{L}) \xrightarrow{\sim} \mathbb{L}$$

*assigns to the vector  $\mathbf{a}_1 \wedge \dots \wedge \mathbf{a}_{n-1}$  the vector  $\mathbf{b} \in \mathbb{L}$ , which is orthogonal to the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_{n-1}$  and whose length is equal to the unoriented volume  $V(\mathbf{a}_1, \dots, \mathbf{a}_{n-1})$ , or more precisely,*

$$\mathbf{b} = V(\mathbf{a}_1, \dots, \mathbf{a}_{n-1})\mathbf{e}, \quad (10.75)$$

*where  $\mathbf{e} \in \mathbb{L}$  is a vector of unit length orthogonal to the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_{n-1}$  and chosen in such a way that the basis  $\mathbf{a}_1, \dots, \mathbf{a}_{n-1}, \mathbf{e}$  is positively oriented.*

The vector  $\mathbf{b}$  determined by the relationship (10.75) is called the *vector product* of the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_{n-1}$  and is denoted by  $[\mathbf{a}_1, \dots, \mathbf{a}_{n-1}]$ . In the case  $n = 3$ , this definition gives us the vector product of two vectors  $[\mathbf{a}_1, \mathbf{a}_2]$  familiar from analytic geometry.

# Chapter 11

## Quadrics

We have encountered a number of types of spaces consisting of points (affine, affine Euclidean, projective). For all of these spaces, an interesting and important question has been the study of *quadrics* contained in such spaces, that is, sets of points with coordinates  $(x_1, \dots, x_n)$  that in some coordinate system satisfy the single equation

$$F(x_1, \dots, x_n) = 0, \quad (11.1)$$

where  $F$  is a second-degree polynomial in the variables  $x_1, \dots, x_n$ . Let us focus our attention on the fact that by the definition of a polynomial, it is possible in general for there to be present in equation (11.1) both first- and second-degree monomials as well as a constant term.

For each of the spaces of the above-mentioned types, a trivial verification shows that the property of a set of points being a quadric does not depend on the choice of coordinate system. Or in other words, a nonsingular affine transformation, motion, or projective transformation (depending on the type of space under consideration) takes a quadric to a quadric.

### 11.1 Quadrics in Projective Space

By the definition given above, a quadric  $Q$  in the projective space  $\mathbb{P}(L)$  is given by equation (11.1) in homogeneous coordinates. However, as we saw in Chap. 9, such an equation is satisfied by the homogeneous coordinates of a point of the projective space  $\mathbb{P}(L)$  only if its left-hand side is homogeneous.

**Definition 11.1** A *quadric* in a projective space  $\mathbb{P}(L)$  is a set  $Q$  consisting of points defined by equation (11.1), where  $F$  is a homogeneous second-degree polynomial, that is, a quadratic form in the coordinates  $x_0, x_1, \dots, x_n$ .

In Sect. 6.2, it was proved that in some coordinate system (that is, in some basis of the space  $L$ ), equation (11.1) is reduced to canonical form

$$\lambda_0 x_0^2 + \lambda_1 x_1^2 + \cdots + \lambda_r x_r^2 = 0,$$

where all the coefficients  $\lambda_i$  are nonzero. Here the number  $r \leq n$  is equal to the rank of the quadratic form  $F$ , and it is the same for every system of coordinates in which the form  $F$  is reduced to canonical form. In the sequel, we shall assume that the quadratic form  $F$  is nonsingular, that is, that  $r = n$ . We shall also call the associated quadric  $Q$  *nonsingular*. The canonical form of its equation can then be written as follows:

$$\alpha_0 x_0^2 + \alpha_1 x_1^2 + \cdots + \alpha_n x_n^2 = 0, \quad (11.2)$$

where all the coefficients  $\alpha_i$  are nonzero. The general case differs from (11.2) only in the omission of terms containing  $x_i$  with  $i = r + 1, \dots, n$ . It is therefore easily reduced to the case of a nonsingular quadric.

We have already encountered the concept of a tangent space to an arbitrary smooth hypersurface (in Chap. 7) or to a projective algebraic variety (in Chap. 9). Now we move on to a consideration of the notion of the tangent space to a quadric.

**Definition 11.2** If  $A$  is a point on the quadric  $Q$  given by equation (11.1), then the *tangent space* to  $Q$  at the point  $A \in Q$  is defined as the projective space  $T_A Q$  given by equation

$$\sum_{i=0}^n \frac{\partial F}{\partial x_i}(A) x_i = 0. \quad (11.3)$$

The tangent space is an important general mathematical concept, and we shall now discuss it in the greatest possible generality. Within the framework of a course in algebra, it is natural to limit ourselves to the case in which  $F$  is a homogeneous polynomial of arbitrary degree  $k > 0$ . Then equation (11.1) defines in the space  $\mathbb{P}(L)$  some *hypersurface*  $X$ , and if not all the partial derivatives  $\frac{\partial F}{\partial x_i}(A)$  are equal to zero, then equation (11.3) gives the *tangent hyperplane* to the hypersurface  $X$  at the point  $A$ . We see that in equation (11.3), on the left-hand side appears the differential  $d_A F(\mathbf{x})$  (see Example 3.86 on p. 130), and since this notion was defined so as to be invariant with respect to the choice of coordinate system, the notion of tangent space is also independent of such a choice. The tangent space to the hypersurface  $X$  at the point  $A$  is denoted by  $T_A X$ .

In the sequel, we shall always assume that quadrics are viewed as lying in spaces over a field  $\mathbb{K}$  of characteristic different from 2 (for example, for definiteness, we may assume that the field  $\mathbb{K}$  is either  $\mathbb{R}$  or  $\mathbb{C}$ ). If  $F(\mathbf{x})$  is a quadratic form, then by the assumptions we have made, we can write it in the form

$$F(\mathbf{x}) = \sum_{i,j=0}^n a_{ij} x_i x_j, \quad (11.4)$$

where the coefficients satisfy  $a_{ij} = a_{ji}$ . In other words,  $F(\mathbf{x}) = \varphi(\mathbf{x}, \mathbf{x})$ , where

$$\varphi(\mathbf{x}, \mathbf{y}) = \sum_{i,j=0}^n a_{ij} x_i y_j \quad (11.5)$$

is a symmetric bilinear form (Theorem 6.6). If the point  $A$  corresponds to the vector  $\mathbf{a}$  with coordinates  $(\alpha_0, \alpha_1, \dots, \alpha_n)$ , then

$$\frac{\partial F}{\partial x_i}(A) = 2 \sum_{j=0}^n a_{ij} \alpha_j,$$

and therefore, equation (11.3) takes the form

$$\sum_{i,j=0}^n a_{ij} \alpha_j x_i = 0,$$

or equivalently,  $\varphi(\mathbf{a}, \mathbf{x}) = 0$ . Thus in this case, the tangent hyperplane at the point  $A$  coincides with the orthogonal complement  $\langle \mathbf{a} \rangle^\perp$  to the vector  $\mathbf{a} \in L$  with respect to the bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$ .

The definition of tangent space (11.3) loses sense if all derivatives  $\frac{\partial F}{\partial x_i}(A)$  are equal to zero:

$$\frac{\partial F}{\partial x_i}(A) = 0, \quad i = 0, 1, \dots, n. \quad (11.6)$$

A point  $A$  of the hypersurface  $X$  given by equation (11.1) for which equalities (11.6) are satisfied is called a *singular* or *critical point*. If a hypersurface has no singular points, then it is said to be *smooth*. When the hypersurface  $X$  is a quadric, that is, the polynomial  $F$  is a quadratic form (11.4), then equations (11.6) assume the form

$$\sum_{j=0}^n a_{ij} \alpha_j = 0, \quad i = 0, 1, \dots, n.$$

Since the point  $A$  is in  $\mathbb{P}(L)$ , it follows that not all of its coordinates  $\alpha_i$  are equal to zero. Thus singular points of a quadric  $Q$  are the nonzero solutions of the system of equations

$$\sum_{j=0}^n a_{ij} x_j = 0, \quad i = 0, 1, \dots, n. \quad (11.7)$$

As was shown in Chap. 2, such solutions exist only if the determinant of the matrix  $(a_{ij})$  is equal to zero, and that is equivalent to saying that the quadric  $Q$  is singular. Thus a nonsingular quadric is the same thing as a smooth quadric.

Let us consider the possible mutual relationships between a quadric  $Q$  and a line  $l$  in projective space  $\mathbb{P}(L)$ . First, let us show that either the line  $l$  has not more than two points in common with the quadric  $Q$ , or else it lies entirely in  $Q$ .



Indeed, if a line  $l$  is not contained entirely in  $Q$ , then one can choose a point  $A \in l$ ,  $A \notin Q$ . Let the line  $l$  correspond to some plane  $L' \subset L$ , that is,  $l = \mathbb{P}(L')$ . If  $A = \langle \mathbf{a} \rangle$ , then  $L' = \langle \mathbf{a}, \mathbf{b} \rangle$ , where the vector  $\mathbf{b} \in L$  is not collinear with the vector  $\mathbf{a}$ . In other words, the plane  $L'$  consists of all vectors of the form  $x\mathbf{a} + y\mathbf{b}$ , where  $x$  and  $y$  range over all possible scalars. The points of intersection of the line  $l$  and plane  $Q$  are found from the equation  $F(x\mathbf{a} + y\mathbf{b}) = 0$ , that is, from the equation

$$\begin{aligned} F(x\mathbf{a} + y\mathbf{b}) &= \varphi(x\mathbf{a} + y\mathbf{b}, x\mathbf{a} + y\mathbf{b}) \\ &= F(\mathbf{a})x^2 + 2\varphi(\mathbf{a}, \mathbf{b})xy + F(\mathbf{b})y^2 = 0 \end{aligned} \quad (11.8)$$

in the variables  $x, y$ . The vectors  $x\mathbf{a} + y\mathbf{b}$  with  $y = 0$  give us a point  $A \notin Q$ . Assuming, therefore, that  $y \neq 0$ , we obtain  $t = x/y$ . Then (11.8) gives us a quadratic equation in the variable  $t$ :

$$F(x\mathbf{a} + y\mathbf{b}) = y^2(F(\mathbf{a})t^2 + 2\varphi(\mathbf{a}, \mathbf{b})t + F(\mathbf{b})) = 0.$$

The condition  $A \notin Q$  has the form  $F(\mathbf{a}) \neq 0$ . Consequently, the leading coefficient of the quadratic trinomial  $F(\mathbf{a})t^2 + 2\varphi(\mathbf{a}, \mathbf{b})t + F(\mathbf{b})$  is nonzero, and therefore, the quadratic trinomial itself is not identically zero and cannot have more than two roots.

Let us now consider the mutual arrangement of  $Q$  and  $l$  if the line  $l$  passes through the point  $A \in Q$ . Then, as in the previous case,  $l$  corresponds to the solutions of the quadratic equation (11.8), in which  $F(\mathbf{a}) = 0$ , since  $A \in Q$ . Thus we obtain the equation

$$F(x\mathbf{a} + y\mathbf{b}) = 2\varphi(\mathbf{a}, \mathbf{b})xy + F(\mathbf{b})y^2 = y(2\varphi(\mathbf{a}, \mathbf{b})x + F(\mathbf{b})y) = 0. \quad (11.9)$$

One solution of equation (11.9) is obvious:  $y = 0$ . It precisely corresponds to the point  $A \in Q$ . This solution is unique if and only if  $\varphi(\mathbf{a}, \mathbf{b}) = 0$ , that is, if  $\mathbf{b} \in T_A Q$ . In the latter case, clearly  $l \subset T_A Q$ , and one says that the line  $l$  is *tangent* to the quadric  $Q$  at the point  $A$ .

Thus there are four possible cases of the relationship between a nonsingular quadric  $Q$  and a line  $l$ :

- (1) The line  $l$  has no points in common with the quadric  $Q$ .
- (2) The line  $l$  has precisely two distinct points in common with the quadric  $Q$ .
- (3) The line  $l$  has exactly one point  $A$  in common with the quadric  $Q$ , which is possible if and only if  $l \subset T_A Q$ .
- (4) The line  $l$  lies entirely in  $Q$ .

Of course, there also exist smooth hypersurfaces defined by equation (11.1) of arbitrary degree  $k \geq 1$ . For example, such a hypersurface is given by the equation  $c_0x_0^k + c_1x_1^k + \cdots + c_nx_n^k = 0$ , where all the  $c_i$  are nonzero. In the sequel, we shall consider only smooth hypersurfaces. For these, the left-hand side of equation (11.3) is a *nonnull* linear form on the vector space  $L$ , and this means that it determines a hyperplane in  $L$  and in  $\mathbb{P}(L)$ .

Let us verify that this hyperplane contains the point  $A$ . This means that if the point  $A$  corresponds to the vector  $\mathbf{a} = (\alpha_0, \alpha_1, \dots, \alpha_n)$ , then

$$\sum_{i=0}^n \frac{\partial F}{\partial x_i}(A) \alpha_i = 0.$$

If the degree of the homogeneous polynomial  $F$  is equal to  $k$ , then by Euler's identity (3.68), we have the equality

$$\sum_{i=0}^n \frac{\partial F}{\partial x_i}(A) \alpha_i = \left( \sum_{i=0}^n \frac{\partial F}{\partial x_i} x_i \right)(A) = kF(A).$$

The value of  $F(A)$  is equal to zero, since the point  $A$  lies on the hypersurface  $X$  given by the equation  $F(A) = 0$ .

Now to switch to a more familiar situation, let us consider an affine subspace of  $\mathbb{P}(L)$ , given by the condition  $x_0 \neq 0$ , and let us introduce in it the inhomogeneous coordinates

$$y_i = x_i/x_0, \quad i = 1, \dots, n. \quad (11.10)$$

Let us assume that the point  $A$  lies in this subset (that is, its coordinate  $\alpha_0$  is nonzero) and let us write equation (11.3) in coordinates  $y_i$ . To do so, we must move from the variables  $x_0, x_1, \dots, x_n$  to the variables  $y_1, \dots, y_n$  and rewrite equation (11.3) accordingly. Here we must set

$$F(x_0, x_1, \dots, x_n) = x_0^k f(y_1, \dots, y_n), \quad (11.11)$$

where  $f(y_1, \dots, y_n)$  is a polynomial of degree  $k \geq 1$ , already not necessarily homogeneous (in contrast to  $F$ ). In accord with formula (11.10), let us denote by  $a_1, \dots, a_n$  the inhomogeneous coordinates of the point  $A$ , that is,

$$a_i = \alpha_i/\alpha_0, \quad i = 1, \dots, n.$$

Using general rules for the calculation of partial derivatives, from the representation (11.11), taking into account (11.10), we obtain the formulas

$$\begin{aligned} \frac{\partial F}{\partial x_0} &= kx_0^{k-1}f + x_0^k \sum_{l=1}^n \frac{\partial f}{\partial y_l} \frac{\partial y_l}{\partial x_0} = kx_0^{k-1}f + x_0^k \sum_{l=1}^n \frac{\partial f}{\partial y_l} \left( -\frac{y_l}{x_0} \right) \\ &= kx_0^{k-1}f - x_0^{k-1} \sum_{l=1}^n \frac{\partial f}{\partial y_l} y_l \end{aligned}$$

and

$$\frac{\partial F}{\partial x_i} = x_0^k \sum_{l=1}^n \frac{\partial f}{\partial y_l} \frac{\partial y_l}{\partial x_i} = x_0^k \sum_{l=1}^n \frac{\partial f}{\partial y_l} \left( x_0^{-1} \frac{\partial x_l}{\partial x_i} \right) = x_0^{k-1} \frac{\partial f}{\partial y_i}, \quad i = 1, \dots, n.$$

Now let us find the values of the derivatives calculated above of the function  $F$  at the point  $A$  with inhomogeneous coordinates  $a_1, \dots, a_n$ . The value of  $F(A)$  is zero, since the point  $A$  lies in the hypersurface  $X$  and  $x_0 \neq 0$ . By virtue of the representation (11.11), we obtain from this that  $f(a_1, \dots, a_n) = 0$ . For brevity, we shall employ the notation  $f(A) = f(a_1, \dots, a_n)$  and  $\frac{\partial f}{\partial y_i}(A) = \frac{\partial f}{\partial y_i}(a_1, \dots, a_n)$ . Thus from the two previous relationships, we obtain

$$\begin{aligned} \frac{\partial F}{\partial x_0}(A) &= -\alpha_0^{k-1} \sum_{i=1}^n \frac{\partial f}{\partial y_i}(A) a_i, \\ \frac{\partial F}{\partial x_i}(A) &= \alpha_0^{k-1} \frac{\partial f}{\partial y_i}(A), \quad i = 1, \dots, n. \end{aligned} \quad (11.12)$$

On substituting expression (11.12) into (11.3), and taking into account (11.10), we obtain the equation

$$\begin{aligned} -\alpha_0^{k-1} \sum_{i=1}^n \frac{\partial f}{\partial y_i}(A) a_i x_0 + \sum_{i=1}^n \left( \alpha_0^{k-1} \frac{\partial f}{\partial y_i}(A) \right) x_i \\ = \alpha_0^{k-1} x_0 \sum_{i=1}^n \frac{\partial f}{\partial y_i}(A) (y_i - a_i) = 0. \end{aligned}$$

Canceling the nonzero common factor  $\alpha_0^{k-1} x_0$ , we finally obtain

$$\sum_{i=1}^n \frac{\partial f}{\partial y_i}(A) (y_i - a_i) = 0. \quad (11.13)$$

This is precisely the equation of the tangent hyperplane  $T_A X$  in inhomogeneous coordinates. In analysis and geometry, it is written in the form (11.13) for a function  $f$  of a much more general class than that of polynomials.

We may now return to the case in which the hypersurface  $X = Q$  is a nonsingular (and therefore smooth) quadric. Then for every point  $A \in Q$ , equation (11.3) determines a hyperplane in  $L$ , that is, some line in the dual space  $L^*$ , and therefore a point belonging to the space  $\mathbb{P}(L^*)$ , which we shall denote by  $\Phi(A)$ . Thus we define the mapping

$$\Phi : Q \rightarrow \mathbb{P}(L^*). \quad (11.14)$$

Our first task consists in determining what the set  $\Phi(Q) \subset \mathbb{P}(L^*)$  in fact is. For this, we express the quadratic form  $F(\mathbf{x})$  in the form  $F(\mathbf{x}) = \varphi(\mathbf{x}, \mathbf{x})$ , where the symmetric bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  has the form (11.5). By Theorem 6.3, we can write  $\varphi(\mathbf{x}, \mathbf{y})$  uniquely as  $\varphi(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathcal{A}(\mathbf{y}))$ , where  $\mathcal{A} : L \rightarrow L^*$  is some linear transformation. From the definitions, it follows that here, the radical of the form  $\varphi$  coincides with the kernel of the linear transformation  $\mathcal{A}$ . Since in the case of a nonsingular form  $F$ , the radical  $\varphi$  is equal to  $(\mathbf{0})$ , it follows that the kernel of  $\mathcal{A}$  is also equal to  $(\mathbf{0})$ . Since  $\dim L = \dim L^*$ , we have by Theorem 3.68 that the linear transformation

$\mathcal{A}$  is an isomorphism, and there is thereby determined a projective transformation  $\mathbb{P}(\mathcal{A}) : \mathbb{P}(L) \rightarrow \mathbb{P}(L^*)$ .

Let us now write down our mapping (11.14) in coordinates. If the quadratic form  $F(x)$  is written in the form (11.4), then

$$\frac{\partial F}{\partial x_i} = 2 \sum_{j=0}^n a_{ij} x_j, \quad i = 0, 1, \dots, n.$$

On the other hand, in some basis  $e_0, e_1, \dots, e_n$  of the space  $L$ , the bilinear form  $\varphi(x, y)$  has the form (11.5), where the vectors  $x$  and  $y$  are given by  $x = x_0 e_0 + \dots + x_n e_n$  and  $y = y_0 e_0 + \dots + y_n e_n$ . From this, it follows that the matrix of the transformation  $\mathcal{A} : L \rightarrow L^*$  in the basis  $e_0, e_1, \dots, e_n$  of the space  $L$  and in the dual basis  $f_0, f_1, \dots, f_n$  of the space  $L^*$  is equal to  $(a_{ij})$ . Therefore, to the quadratic form  $F(x)$  is associated the isomorphism  $\mathcal{A} : L \rightarrow L^*$ , and the mapping (11.14) that we constructed coincides with the restriction of the projective transformation  $\mathbb{P}(\mathcal{A}) : \mathbb{P}(L) \rightarrow \mathbb{P}(L^*)$  to  $Q$ , that is,  $\Phi(Q) = \mathbb{P}(\mathcal{A})(Q)$ .

From this arises an unexpected consequence: since the transformation  $\mathbb{P}(\mathcal{A})$  is a bijection, the transformation (11.14) is also a bijection. In other words, the tangent hyperplanes to the nonsingular quadric  $Q$  at distinct points  $A, B \in Q$  are distinct. Thus we obtain the following result.

**Lemma 11.3** *The same hyperplane cannot coincide with the tangent hyperplanes to a nonsingular quadric  $Q$  at two distinct points.*

This means that in writing a hyperplane of the space  $\mathbb{P}(L)$  in the form  $T_A Q$ , we may omit the point  $A$ . And in the case of a nonsingular quadric  $Q$ , it makes sense to say that the *hyperplane is tangent to the quadric*, and moreover, the point of tangency  $A \in Q$  is uniquely determined.

Let us now consider more concretely what the set  $\Phi(Q)$  looks like. We shall show that it is also a nonsingular quadric, that is, in some (and therefore in any) basis of the space  $L^*$  determined by the equation  $q(x) = 0$ , where  $q$  is a nonsingular quadratic form.

We saw above that there is an isomorphism  $\mathcal{A} : L \xrightarrow{\sim} L^*$  that bijectively maps  $Q$  to  $\Phi(Q)$ . Therefore, there exists as well an inverse transformation  $\mathcal{A}^{-1} : L^* \xrightarrow{\sim} L$ , which is also an isomorphism. Then the condition  $y \in \Phi(Q)$  is equivalent to  $\mathcal{A}^{-1}(y) \in Q$ . Let us choose an arbitrary basis

$$f_0, f_1, \dots, f_n \tag{11.15}$$

in the space  $L^*$ . The isomorphism  $\mathcal{A}^{-1} : L^* \xrightarrow{\sim} L$  carries this basis to the basis

$$\mathcal{A}^{-1}(f_0), \mathcal{A}^{-1}(f_1), \dots, \mathcal{A}^{-1}(f_n) \tag{11.16}$$

of the space  $L$ . Here obviously the coordinates of the vector  $\mathcal{A}^{-1}(y)$  in the basis (11.16) coincide with the coordinates of the vector  $y$  in the basis (11.15). As we

saw above, the condition  $\mathcal{A}^{-1}(\mathbf{y}) \in Q$  is equivalent to the relationship

$$F(\alpha_0, \alpha_1, \dots, \alpha_n) = 0, \quad (11.17)$$

where  $F$  is a nonsingular quadratic form, and  $(\alpha_0, \alpha_1, \dots, \alpha_n)$  are the coordinates of the vector  $\mathcal{A}^{-1}(\mathbf{y})$  in some basis of the space  $L$ , for instance, in the basis (11.16). This means that the condition  $\mathbf{y} \in \Phi(Q)$  can be expressed by the same relationship (11.17). Thus we have proved the following statement.

**Theorem 11.4** *If  $Q$  is a nonsingular quadric in the space  $\mathbb{P}(L)$ , then the set of tangent hyperplanes to it forms a nonsingular quadric in the space  $\mathbb{P}(L^*)$ .*

Repeating verbatim the arguments presented in Sect. 9.1, we may extend the duality principle formulated there. Namely, we can add to it some additional notions that are dual to each other that can be interchanged so that the general assertion formulated on p. 326 remains valid:

$$\begin{array}{l} \text{nonsingular quadric in } \mathbb{P}(L) \\ \text{point in a nonsingular quadric} \end{array} \quad \left\| \quad \begin{array}{l} \text{nonsingular quadric in } \mathbb{P}(L^*) \\ \text{hyperplane tangent to a nonsingular quadric} \end{array} \right.$$

This (seemingly small) extension of the duality principle leads to completely unexpected results. By way of an example, we shall introduce two famous theorems that are duals of each other, that is, equivalent on the basis of the duality principle. Yet the second of them was published 150 years after the first. These theorems relate to quadrics in two-dimensional projective space, that is, in the projective plane. In this case, a quadric is called a *conic*.<sup>1</sup>

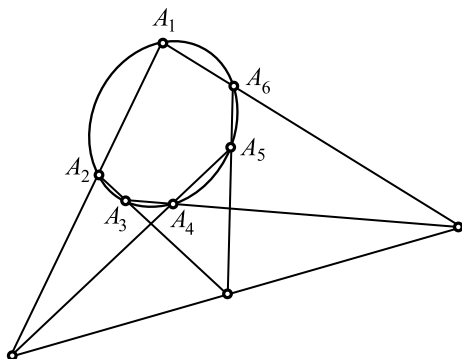
In the sequel, we shall use the following terminology. Let  $Q$  be a nonsingular conic, and let  $A_1, \dots, A_6$  be six distinct points of  $Q$ . This ordered (that is, their order is significant) collection of points is called a *hexagon inscribed in the conic  $Q$* . For two distinct points  $A$  and  $B$  of the projective plane, their projective cover (that is, the line passing through them) is denoted by  $\overline{AB}$  (cf. the definition on p. 325). The six lines  $\overline{A_1A_2}, \overline{A_2A_3}, \dots, \overline{A_5A_6}, \overline{A_6A_1}$  are called the *sides* of the hexagon.<sup>2</sup> Here the following pairs of sides will be called *opposite sides*:  $\overline{A_1A_2}$  and  $\overline{A_4A_5}$ ,  $\overline{A_2A_3}$  and  $\overline{A_5A_6}$ ,  $\overline{A_3A_4}$  and  $\overline{A_6A_1}$ .

**Theorem 11.5** (Pascal's theorem) *Pairs of opposite sides of an arbitrary hexagon inscribed in a nonsingular cone intersect in three collinear points. See Fig. 11.1.*

<sup>1</sup>A clarification of this term, that is, an explanation of what this has to do with a cone, will be given somewhat later.

<sup>2</sup>Here we move away somewhat from the intuition of elementary geometry, where by a side we mean not the entire line passing through two points, but only the segment connecting them. This extended notion of a side is necessary if we wish to include the case of an arbitrary field  $\mathbb{K}$ , for instance,  $\mathbb{K} = \mathbb{C}$ .

**Fig. 11.1** Hexagon inscribed in a conic



Before formulating the dual theorem to Pascal's theorem, let us make a few comments.

With the selection of a homogeneous system of coordinates  $(x_0 : x_1 : x_2)$  in the projective plane, the equation of the conic  $Q$  can be written in the form

$$F(x_0 : x_1 : x_2) = a_1 x_0^2 + a_2 x_0 x_1 + a_3 x_0 x_2 + a_4 x_1^2 + a_5 x_1 x_2 + a_6 x_2^2 = 0.$$

There are six coefficients on the right-hand side of this equation. If we have  $k$  points  $A_1, \dots, A_k$ , then the condition of their belonging to the conic  $Q$  reduces to the relationships

$$F(A_i) = 0, \quad i = 1, \dots, k, \quad (11.18)$$

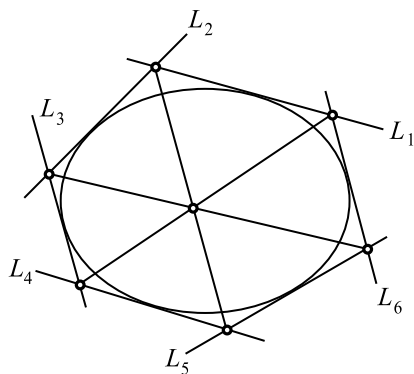
which yield a system consisting of  $k$  linear homogeneous equations in the six unknowns  $a_1, \dots, a_6$ . We must find a nontrivial solution to this system. If we have  $k = 6$ , then this question falls under Corollary 2.13 as a special case (and this explains our interest in *hexagons* inscribed in a conic). By this corollary, we have still to verify that the determinant of the system (11.18) for  $k = 6$  is equal to zero. It is Pascal's theorem that gives a geometric interpretation of this condition.

It is not difficult to show that it gives *necessary and sufficient* conditions for six points  $A_1, \dots, A_6$  to lie on some conic if we restrict ourselves, first of all, to nonsingular conics, and secondly, to such collections of six points that no three of them are collinear (this is proved in any sufficiently rigorous course in analytic geometry).

Now let us formulate the dual theorem to Pascal's theorem. Here six distinct lines  $L_1, \dots, L_6$  tangent to a conic  $Q$  will be called a *hexagon circumscribed about the conic*. Points  $L_1 \cap L_2, L_2 \cap L_3, L_3 \cap L_4, L_4 \cap L_5, L_5 \cap L_6$ , and  $L_6 \cap L_1$  are called the *vertices* of the hexagon. Here the following pairs of vertices will be called *opposite*:  $L_1 \cap L_2$  and  $L_4 \cap L_5$ ,  $L_2 \cap L_3$  and  $L_5 \cap L_6$ ,  $L_3 \cap L_4$  and  $L_6 \cap L_1$ .

**Theorem 11.6** (Brianchon's theorem) *The lines connecting opposite vertices of an arbitrary hexagon circumscribed about a nonsingular conic intersect at a common point. See Fig. 11.2.*

**Fig. 11.2** Hexagon circumscribed about a conic



It is obvious that Brianchon's theorem is obtained from Pascal's theorem if we replace in it all the concepts by their duals according to the rules given above. Thus by virtue of the general duality principle, Brianchon's theorem follows from Pascal's theorem. Pascal's theorem itself can be proved easily, but we will not present a proof, since its logic is connected with another area, namely algebraic geometry.<sup>3</sup> Here it is of interest to observe only that the duality principle makes it possible to obtain certain results from others that appear at first glance to be entirely unrelated. Indeed, Pascal proved his theorem in the seventeenth century (when he was 16 years old), while Brianchon proved his theorem in the nineteenth century, more than 150 years later. And moreover, Brianchon used entirely different arguments (the general duality principle was not yet understood at the time).

## 11.2 Quadrics in Complex Projective Space

Let us now consider the projective space  $\mathbb{P}(L)$ , where  $L$  is a complex vector space, and as before, let us limit ourselves to the case of nonsingular quadrics. As we saw in Sect. 6.3 (formula (6.27)), a nonsingular quadratic form in a complex space has the canonical form  $x_0^2 + x_1^2 + \cdots + x_n^2$ . This means that in some coordinate system, the equation of a nonsingular quadric can be written as

$$x_0^2 + x_1^2 + \cdots + x_n^2 = 0, \quad (11.19)$$

that is, every nonsingular quadric can be transformed into the quadric (11.19) by some projective transformation. In other words, in a complex projective space there exists (defined up to a projective transformation) *only one* nonsingular quadric (11.19). It is this quadric that we shall now investigate.

In view of what we have said above, it suffices to consider any one arbitrary nonsingular quadric on the projective space  $\mathbb{P}(L)$  of a given dimension. For example,

<sup>3</sup>Such a proof can be found, for example, in the book *Algebraic Curves*, by Robert Walker (Springer, 1978).

we may choose the quadric given by the equation  $F(\mathbf{x}) = 0$ , where the matrix of the quadratic form  $F(\mathbf{x})$  has the form

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{pmatrix}. \quad (11.20)$$

A simple calculation shows that the determinant of the matrix (11.20) is equal to  $+1$  or  $-1$ , that is, it is nonzero.

A fundamental topic that we shall study in this and the following sections is projective subspaces contained in a quadric. Let the quadric  $Q$  be given by the equation  $F(\mathbf{x}) = 0$ , where  $\mathbf{x} \in L$ , and let a projective subspace have the form  $\mathbb{P}(L')$ , where  $L'$  is a subspace of the vector space  $L$ . Then the projective subspace  $\mathbb{P}(L')$  is contained in  $Q$  if and only if  $F(\mathbf{x}) = 0$  for all vectors  $\mathbf{x} \in L'$ .

**Definition 11.7** A subspace  $L' \subset L$  is said to be *isotropic* with respect to a quadratic form  $F$  if  $F(\mathbf{x}) = 0$  for all vectors  $\mathbf{x} \in L'$ .

Let  $\varphi$  be the symmetric bilinear form associated with the quadratic form  $F$ , according to Theorem 6.6. Then by virtue of (6.14), we see that  $\varphi(\mathbf{x}, \mathbf{y}) = 0$  for all vectors  $\mathbf{x}, \mathbf{y} \in L'$ . Therefore, we shall also say that the subspace  $L' \subset L$  is isotropic with respect to the bilinear form  $\varphi$ .

We have already encountered the simplest example of isotropic subspaces, in Sect. 7.7 in our study of pseudo-Euclidean spaces. There we encountered lightlike (also called isotropic) vectors on which a quadratic form ( $\mathbf{x}^2$ ) defining a pseudo-Euclidean space becomes zero. Every nonnull lightlike vector  $\mathbf{e}$  clearly determines a one-dimensional subspace  $\langle \mathbf{e} \rangle$ .

The basic technique that will be used in this and the following sections consists in how to reformulate our questions about subspaces contained in a quadric  $F(\mathbf{x}) = 0$  in terms of a vector space  $L$ , a symmetric bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  defined on  $L$  and corresponding to the quadratic form  $F(\mathbf{x})$ , and subspaces isotropic with respect to  $F$  and  $\varphi$ . Then everything is determined almost trivially on the basis of the simplest properties of linear and bilinear forms.

**Theorem 11.8** *The dimension of an arbitrary isotropic subspace  $L' \subset L$  relative to an arbitrary nonsingular quadratic form  $F$  does not exceed half of  $\dim L$ .*

*Proof* Let us consider  $(L')^\perp$ , the orthogonal complement of the subspace  $L' \subset L$  with respect to the bilinear form  $\varphi(\mathbf{u}, \mathbf{v})$  associated with  $F(\mathbf{x})$ . The quadratic form  $F(\mathbf{x})$  and bilinear form  $\varphi(\mathbf{u}, \mathbf{v})$  are nonsingular. Therefore, we have relationship (7.75), from which follows the equality  $\dim(L')^\perp = \dim L - \dim L'$ .



That the space  $L'$  is isotropic means that  $L' \subset (L')^\perp$ . From this we obtain the inequality

$$\dim L' \leq \dim (L')^\perp = \dim L - \dim L',$$

from which it follows that  $\dim L' \leq \frac{1}{2} \dim L$ , as asserted in the theorem.  $\square$

In the sequel, we shall limit our study of isotropic subspaces to those of the greatest possible dimension, namely  $\frac{1}{2} \dim L$  when the number  $\dim L$  is even and  $\frac{1}{2}(\dim L - 1)$  when it is odd. The general case  $\dim L' \leq \frac{1}{2} \dim L$  is easily reduced to this limiting case and is studied completely analogously.

Let us consider some of the simplest cases, known from analytic geometry.

*Example 11.9* The simplest case of all is  $\dim L = 2$ , and therefore,  $\dim \mathbb{P}(L) = 1$ . In coordinates  $(x_0 : x_1)$ , the quadratic form with matrix (11.20) has the form  $x_0 x_1$ . Clearly, the quadric  $x_0 x_1 = 0$  consists of two points  $(0 : 1)$  and  $(1 : 0)$ , corresponding to the vectors  $e_1 = (0, 1)$  and  $e_2 = (1, 0)$  in the plane  $L$ . Each of the two points determines an isotropic subspace  $L'_i = \langle e_i \rangle$ .

*Example 11.10* Next in complexity is the case  $\dim L = 3$ , and correspondingly,  $\dim \mathbb{P}(L) = 2$ . In this case, we are dealing with quadrics in the projective plane; their points determine one-dimensional isotropic subspaces in  $L$  that therefore form a continuous family. (If the equation of the quadric is  $F(x_0, x_1, x_2) = 0$ , then in the space  $L$ , it determines a quadratic cone whose generatrices are isotropic subspaces.)

*Example 11.11* The following case corresponds to  $\dim L = 4$  and  $\dim \mathbb{P}(L) = 3$ . These are quadrics in three-dimensional projective space. For isotropic subspaces  $L' \subset L$ , Theorem 11.8 gives  $\dim L' \leq 2$ . Isotropic subspaces of maximal dimension are obtained for  $\dim L' = 2$ , that is,  $\dim \mathbb{P}(L') = 1$ . These are projective lines lying on the quadric. In coordinates  $(x_0 : x_1 : y_0 : y_1)$ , the quadratic form with matrix (11.20) gives the equation

$$x_0 y_0 + x_1 y_1 = 0. \quad (11.21)$$

We must find all two-dimensional isotropic subspaces  $L' \subset L$ . Let a basis of the two-dimensional subspace  $L'$  consist of vectors  $e = (a_0, a_1, b_0, b_1)$  and  $e' = (a'_0, a'_1, b'_0, b'_1)$ . Then the fact that  $L'$  is isotropic is expressed, in view of formula (11.21), by the relationship

$$(a_0 u + a'_0 v)(b_0 u + b'_0 v) + (a_1 u + a'_1 v)(b_1 u + b'_1 v) = 0, \quad (11.22)$$

which is satisfied identically for all  $u$  and  $v$ . The left-hand side of equation (11.22) represents a quadratic form in the variables  $u$  and  $v$ , which can be identically equal to zero only in the case that all its coefficients are equal to zero. Removing parentheses in (11.22), we obtain

$$\begin{aligned} a_0 b_0 + a_1 b_1 &= 0, & a_0 b'_0 + a'_0 b_0 + a_1 b'_1 + a'_1 b_1 &= 0, \\ a'_0 b'_0 + a'_1 b'_1 &= 0. \end{aligned} \quad (11.23)$$

The first equation from (11.23) means that the rows  $(a_0, a_1)$  and  $(b_1, -b_0)$  are proportional. Since they cannot both be equal to zero simultaneously (then all coordinates of the basis vector  $e$  would be equal to zero, which is impossible), it follows that one of them is the product of the other and some (uniquely determined) scalar  $\beta$ . For definiteness, let  $a_0 = \beta b_1$ ,  $a_1 = -\beta b_0$  (the case  $b_1 = \beta a_0$ ,  $b_0 = -\beta a_1$  is considered analogously). In just the same way, from the third equation of (11.23), we obtain that  $a'_0 = \gamma b'_1$ ,  $a'_1 = -\gamma b'_0$  with some scalar  $\gamma$ . Substituting the relationships

$$a_0 = \beta b_1, \quad a_1 = -\beta b_0, \quad a'_0 = \gamma b'_1, \quad a'_1 = -\gamma b'_0 \quad (11.24)$$

into the second equation of (11.23), we obtain the equality  $(\beta - \gamma)(b'_0 b_1 - b_0 b'_1) = 0$ . Therefore, either  $b'_0 b_1 - b_0 b'_1 = 0$  or  $\gamma = \beta$ .

In the first case, from the equality  $b'_0 b_1 - b_0 b'_1 = 0$  it follows that the rows  $(b_0, b'_0)$  and  $(b_1, b'_1)$  are proportional, and we obtain the relationships  $b_1 = -\alpha b_0$  and  $b'_1 = -\alpha b'_0$  with some scalar  $\alpha$  (the case  $b_0 = -\alpha b_1$  and  $b'_0 = -\alpha b'_1$  is considered similarly). Let us assume that  $b_1$  and  $b'_1$  are not both equal to zero. Then  $\alpha \neq 0$ , and taking into account the relationships (11.24), we obtain

$$\begin{aligned} a_0 u + a'_0 v &= a_0 u + a'_0 v = \beta b_1 u + \gamma b'_1 v = -\alpha(\beta b_0 u + \gamma b'_0 v) = \alpha(a_1 u + a'_1 v), \\ b_0 u + b'_0 v &= -\alpha^{-1}(b_1 u + b'_1 v). \end{aligned}$$

In the second case, let us suppose that  $a_0$  and  $a_1$  are not both equal to zero. Then  $\beta \neq 0$ , and taking into account relationship (11.24), we obtain

$$\begin{aligned} a_0 u + a'_0 v &= a_0 u + a'_0 v = \beta(b_1 u + b'_1 v), \\ b_0 u + b'_0 v &= -\beta^{-1}(a_1 u + a'_1 v). \end{aligned}$$

Thus with the assumptions made for an arbitrary vector subspace  $L'$  with coordinates  $(x_0, y_0, x_1, y_1)$ , we have either

$$x_0 = \alpha x_1, \quad y_0 = -\alpha^{-1} y_1 \quad (11.25)$$

or

$$x_0 = \beta y_1, \quad y_0 = -\beta^{-1} x_1, \quad (11.26)$$

where  $\alpha$  and  $\beta$  are certain nonzero scalars.

In order to consider the excluded cases, namely  $\alpha = 0$  ( $b_1 = b'_1 = 0$ ) and  $\beta = 0$  ( $a_0 = a_1 = 0$ ), let us introduce points  $(a : b) \in \mathbb{P}^1$  and  $(c : d) \in \mathbb{P}^1$ , that is, pairs of numbers that are not simultaneously equal to zero, and let us consider them as defined up to multiplication by one and the same nonzero scalar. Then as is easily verified, a homogeneous representation of relationships (11.25) and (11.26) that also includes both previously excluded cases will have the form

$$ax_0 = bx_1, \quad by_0 = -ay_1 \quad (11.27)$$

and

$$cx_0 = dy_1, \quad dy_0 = -cx_1 \quad (11.28)$$

respectively. Indeed, equality (11.25) is obtained from (11.27) for  $a = 1$  and  $b = \alpha$ , while (11.26) is obtained from (11.28) for  $c = 1$  and  $d = \beta$ .

Relationships (11.27) give the isotropic plane  $L' \subset L$  or the line  $\mathbb{P}(L')$  in  $\mathbb{P}(L)$ , which belongs to the quadric (11.21). It is determined by the point  $(a : b) \in \mathbb{P}^1$ . Thus we obtain *one family* of lines. Similarly, relationships (11.28) determine a *second family* of lines. Together, they give all the lines contained in our quadric (called a *hyperboloid of one sheet*). These lines are called the *rectilinear generatrices* of the hyperboloid.

On the basis of the formulas we have written down, it is easy to verify some properties known from analytic geometry: two distinct lines from one family of rectilinear generatrices do not intersect, while two lines from different families do intersect (at a single point). For every point of the hyperboloid, there is a line from each of the two families that passes through it.

In the following section, we shall consider the general case of projective subspaces of maximum possible dimension on a nonsingular quadric of arbitrary dimension in complex projective space.

### 11.3 Isotropic Subspaces

Let  $Q$  be a nonsingular quadric in a complex projective space  $\mathbb{P}(L)$  given by the equation  $F(x) = 0$ , where  $F(x)$  is a nonsingular quadratic form on the space  $L$ . In analogy to what we discussed in the previous section, we shall study  $m$ -dimensional subspaces  $L' \subset L$  that are isotropic with respect to  $F$ , assuming that  $\dim L = 2m$  if  $\dim L$  is even, and  $\dim L = 2m + 1$  if  $\dim L$  is odd.

The special cases that we studied in the preceding section show that isotropic subspaces look different for different values of  $\dim L$ . Thus for  $\dim L = 3$ , we found *one* family of isotropic subspaces, continuously parameterized by the points of the quadric  $Q$ . For  $\dim L = 2$  or  $4$ , we found *two* such families. This leads to the idea that the number of continuously parameterized families of isotropic subspaces on a quadric depends on the parity of the number  $\dim L$ . As we shall now see, such is indeed the case.

The cases of even and odd dimension will be treated separately.

*Case 1.* Let us assume that  $\dim L = 2m$ . Consequently, we are interested in isotropic subspaces  $M \subset L$  of dimension  $m$ . (This is the most interesting case, since here we shall see how the families of lines on a hyperbola of one sheet are generalized.)

**Theorem 11.12** *For every  $m$ -dimensional isotropic subspace  $M \subset L$ , there exists another  $m$ -dimensional isotropic subspace  $N \subset L$  such that*

$$L = M \oplus N. \quad (11.29)$$

*Proof* Our proof is by induction on the number  $m$ . For  $m = 0$ , the statement of the theorem is vacuously true.

Let us assume now that  $m > 0$ , and let us consider an arbitrary nonnull vector  $\mathbf{e} \in M$ . Let  $\varphi(\mathbf{x}, \mathbf{y})$  be the symmetric bilinear form associated with the quadratic form  $F(\mathbf{x})$ . Since the subspace  $M$  is isotropic, it follows that  $\varphi(\mathbf{e}, \mathbf{e}) = 0$ . In view of the nonsingularity of  $F(\mathbf{x})$ , the bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  is likewise nonsingular, and therefore, its radical is equal to  $(\mathbf{0})$ . Then the linear function  $\varphi(\mathbf{e}, \mathbf{x})$  of a vector  $\mathbf{x} \in L$  is not identically equal to zero (otherwise, the vector  $\mathbf{e}$  would be in the radical of  $\varphi(\mathbf{x}, \mathbf{y})$ , which is equal to  $(\mathbf{0})$ ).

Let  $\mathbf{f} \in L$  be a vector such that  $\varphi(\mathbf{e}, \mathbf{f}) \neq 0$ . Clearly, the vectors  $\mathbf{e}, \mathbf{f}$  are linearly independent. Let us consider the plane  $W = \langle \mathbf{e}, \mathbf{f} \rangle$  and denote by  $\varphi'$  the restriction of the bilinear form  $\varphi$  to  $W$ . In the basis  $\mathbf{e}, \mathbf{f}$ , the matrix of the bilinear form  $\varphi'$  has the form

$$\Phi' = \begin{pmatrix} 0 & \varphi(\mathbf{e}, \mathbf{f}) \\ \varphi(\mathbf{e}, \mathbf{f}) & \varphi(\mathbf{f}, \mathbf{f}) \end{pmatrix}, \quad \varphi(\mathbf{e}, \mathbf{f}) \neq 0.$$

It is obvious that  $|\Phi'| = -\varphi(\mathbf{e}, \mathbf{f})^2 \neq 0$ , and therefore, the bilinear form  $\varphi'$  is nonsingular.

Let us define the vector

$$\mathbf{g} = \mathbf{f} - \frac{\varphi(\mathbf{f}, \mathbf{f})}{2\varphi(\mathbf{e}, \mathbf{f})}\mathbf{e}.$$

Then as is easily verified,  $\varphi(\mathbf{g}, \mathbf{g}) = 0$ ,  $\varphi(\mathbf{e}, \mathbf{g}) = \varphi(\mathbf{e}, \mathbf{f}) \neq 0$ , and the vectors  $\mathbf{e}, \mathbf{g}$  are linearly independent, that is,  $W = \langle \mathbf{e}, \mathbf{g} \rangle$ . In the basis  $\mathbf{e}, \mathbf{g}$ , the matrix of the bilinear form  $\varphi'$  has the form

$$\Phi'' = \begin{pmatrix} 0 & \varphi(\mathbf{e}, \mathbf{g}) \\ \varphi(\mathbf{e}, \mathbf{g}) & 0 \end{pmatrix}.$$

As a result of the nondegeneracy of the bilinear form  $\varphi'$ , we have by Theorem 6.9 the decomposition

$$L = W \oplus L_1, \quad L_1 = W_{\varphi}^{\perp}, \quad (11.30)$$

where  $\dim L_1 = 2m - 2$ . Let us set  $M_1 = L_1 \cap M$  and show that  $M_1$  is a subspace of dimension  $m - 1$  isotropic with respect to the restriction of the bilinear form  $\varphi$  to  $L_1$ .

By construction, the subspace  $M_1$  consists of the vectors  $\mathbf{x} \in M$  such that  $\varphi(\mathbf{x}, \mathbf{e}) = 0$  and  $\varphi(\mathbf{x}, \mathbf{g}) = 0$ . But the first equality holds in general for all  $\mathbf{x} \in M$ , since  $\mathbf{e} \in M$  and  $M$  is isotropic with respect to  $\varphi$ . Thus in the definition of the subspace  $M_1$ , there remains only the second equality, which means that  $M_1 \subset M$  is determined by what is sent to zero by the linear function  $f(\mathbf{x}) = \varphi(\mathbf{x}, \mathbf{g})$ , which is not identically equal to zero (since  $f(\mathbf{e}) = \varphi(\mathbf{e}, \mathbf{g}) \neq 0$ ). Therefore,  $\dim M_1 = \dim M - 1 = m - 1$ .

Thus  $M_1$  is a subspace of  $L_1$  of half the dimension of  $L_1$ , defined by formula (11.30), and we can apply the induction hypothesis to it to obtain the decomposition

$$L_1 = M_1 \oplus N_1, \quad (11.31)$$

where  $N_1 \subset L_1$  is some other  $(m-1)$ -dimensional isotropic subspace.

Let us note that  $M = \langle e \rangle \oplus M_1$  and let us set  $N = \langle g \rangle \oplus N_1$ . Since the subspace  $N_1$  is isotropic in  $L_1$ , the subspace  $N$  is isotropic in  $L$ , and taking into account that  $\varphi(g, g) = 0$ , we have for all vectors  $x \in N_1$  the equality  $\varphi(g, x) = 0$ . Formulas (11.30) and (11.31) together give the decomposition

$$L = \langle e \rangle \oplus \langle g \rangle \oplus M_1 \oplus N_1 = M \oplus N,$$

which is what was to be proved.  $\square$

In the terminology of Theorem 11.12, an arbitrary vector  $z \in N$  determines a linear function  $f(x) = \varphi(z, x)$  on the vector space  $L$ , that is, an element of the dual space  $L^*$ . The restriction of this function to the subspace  $M \subset L$  is obviously a linear function on  $M$ , that is, an element of the space  $M^*$ . This defines the mapping  $\mathcal{F} : N \rightarrow M^*$ . A trivial verification shows that  $\mathcal{F}$  is a linear transformation.

The decomposition (11.29) established by Theorem 11.12 has an interesting consequence.

**Lemma 11.13** *The linear transformation  $\mathcal{F} : N \rightarrow M^*$  constructed above is an isomorphism.*

*Proof* Let us determine the kernel of the transformation  $\mathcal{F} : N \rightarrow M^*$ . Let us assume that  $\mathcal{F}(z_0) = 0$  for some  $z_0 \in N$ , that is,  $\varphi(z_0, y) = 0$  for all vectors  $y \in M$ . But by Theorem 11.12, every vector  $x \in L$  can be represented in the form  $x = y + z$ , where  $y \in M$  and  $z \in N$ . Thus

$$\varphi(z_0, x) = \varphi(z_0, y) + \varphi(z_0, z) = \varphi(z_0, z) = 0,$$

since both vectors  $z$  and  $z_0$  belong to the isotropic subspace  $N$ . From the nonsingularity of the bilinear form  $\varphi$ , it then follows that  $z_0 = 0$ , that is, the kernel of  $\mathcal{F}$  consists of only the null vector. Since  $\dim M = \dim N$ , we have by Theorem 3.68 that the linear transformation  $\mathcal{F}$  is an isomorphism.  $\square$

Let  $e_1, \dots, e_m$  be some basis in  $M$ , and  $f_1, \dots, f_m$  the dual basis in  $M^*$ . The isomorphism  $\mathcal{F}$  that we constructed creates a correspondence between this dual basis and a certain basis  $g_1, \dots, g_m$  in the space  $N$  according to the formula  $\mathcal{F}(g_i) = f_i$ . From decomposition (11.29) established in Theorem 11.12, it follows that vectors  $e_1, \dots, e_m, g_1, \dots, g_m$  form a basis in  $L$ . In this basis, the bilinear form  $\varphi$  has the simplest possible matrix  $\Phi$ . Indeed, recalling the definitions of concepts that we have used, we obtain that

$$\Phi = \begin{pmatrix} 0 & E \\ E & 0 \end{pmatrix}, \quad (11.32)$$

where  $E$  and  $0$  are the identity and zero matrices of order  $m$ . For the corresponding quadratic form  $F$  and vector

$$\mathbf{x} = x_1 \mathbf{e}_1 + \cdots + x_m \mathbf{e}_m + x_{m+1} \mathbf{g}_1 + \cdots + x_{2m} \mathbf{g}_m,$$

we obtain

$$F(\mathbf{x}) = \sum_{i=1}^m x_i x_{m+i}. \quad (11.33)$$

Conversely, if in some basis  $\mathbf{e}_1, \dots, \mathbf{e}_{2m}$  of the vector space  $L$ , the bilinear form  $\varphi$  has matrix (11.32), then the space  $L$  can be represented in the form

$$L = M \oplus N, \quad M = \langle \mathbf{e}_1, \dots, \mathbf{e}_m \rangle, N = \langle \mathbf{e}_{m+1}, \dots, \mathbf{e}_{2m} \rangle,$$

in accordance with Theorem 11.12. Let us recall that in our case (in a complex projective space), all nonsingular bilinear forms are equivalent, and therefore, every nonsingular bilinear form  $\varphi$  has matrix (11.32) in some basis. In particular, we see that in the  $2m$ -dimensional space  $L$ , there exists an  $m$ -dimensional isotropic subspace  $M$ .

In order to generalize known results from analytic geometry for  $m = 2$  to the case of arbitrary  $m$  (see Example 11.11), we shall provide several definitions that naturally generalize some concepts about Euclidean spaces familiar to us from Chap. 7.

**Definition 11.14** Let  $\varphi(\mathbf{x}, \mathbf{y})$  be a nonsingular symmetric bilinear form in the space  $L$  of arbitrary dimension. A linear transformation  $\mathcal{U} : L \rightarrow L$  is said to be *orthogonal* with respect to  $\varphi$  if

$$\varphi(\mathcal{U}(\mathbf{x}), \mathcal{U}(\mathbf{y})) = \varphi(\mathbf{x}, \mathbf{y}) \quad (11.34)$$

for all vectors  $\mathbf{x}, \mathbf{y} \in L$ .

This definition generalizes the notion of orthogonal transformation of a Euclidean space and Lorentz transformation of a pseudo-Euclidean space. Similarly, we shall call a basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of a space  $L$  *orthonormal* with respect to a bilinear form  $\varphi$  if  $\varphi(\mathbf{e}_i, \mathbf{e}_i) = 1$  and  $\varphi(\mathbf{e}_i, \mathbf{e}_j) = 0$  for all  $i \neq j$ . Every orthogonal transformation takes an orthonormal basis into an orthonormal basis, and for any two orthonormal bases, there exists a unique orthogonal transformation taking the first of them to the second. The proofs of these assertions coincide word for word with the analogous assertions from Section 7.2, since there we nowhere used the positive definiteness of the bilinear form  $(\mathbf{x}, \mathbf{y})$ , but only its nonsingularity.

The condition (11.34) can be expressed in matrix form. Let the bilinear form  $\varphi$  have matrix  $\Phi$  in some basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the space  $L$ . Then the transformation  $\mathcal{U} : L \rightarrow L$  will be orthogonal with respect to  $\varphi$  if and only if its matrix  $U$  in this basis satisfies the relationship

$$U^* \Phi U = \Phi. \quad (11.35)$$

This is proved just as was the analogous equality (7.18) for orthogonal transformations of Euclidean spaces, and (7.18) is a special case of formula (11.35) for  $\Phi = E$ .

It follows from formula (11.35) that  $|U^*| \cdot |\Phi| \cdot |U| = |\Phi|$ , and taking into account the nonsingularity of the form  $\varphi$  ( $|\Phi| \neq 0$ ), that  $|U^*| \cdot |U| = 1$ , that is,  $|U|^2 = 1$ . From this we finally obtain the equality  $|U| = \pm 1$ , in which  $|U|$  can be replaced by  $|\mathcal{U}|$ , since the determinant of a linear transformation does not depend on the choice of basis in the space, and consequently, coincides with the determinant of the matrix of this transformation.

The equality  $|\mathcal{U}| = \pm 1$  generalizes a well-known property of orthogonal transformations of a Euclidean space and provides justification for an analogous definition.

**Definition 11.15** A linear transformation  $\mathcal{U} : L \rightarrow L$  orthogonal with respect to a symmetric bilinear form  $\varphi$  is said to be *proper* if  $|\mathcal{U}| = 1$  and *improper* if  $|\mathcal{U}| = -1$ .

It follows at once from Theorem 2.54 on the determinant of the product of matrices that proper and improper transformations multiply just like the numbers  $+1$  and  $-1$ . Similarly, the transformation  $\mathcal{U}^{-1}$  corresponds to the same type (of proper or improper orthogonal transformation) as  $\mathcal{U}$ .

The concepts that we have introduced can be applied to the theory of isotropic subspaces on the basis of the following result.

**Theorem 11.16** For any two  $m$ -dimensional isotropic subspaces  $M$  and  $M'$  of a  $2m$ -dimensional space  $L$ , there exists an orthogonal transformation  $\mathcal{U} : L \rightarrow L$  taking one of the subspaces to the other.

*Proof* Since Theorem 11.12 can be applied to each of the subspaces  $M$  and  $M'$ , there exist  $m$ -dimensional isotropic subspaces  $N$  and  $N'$  such that

$$L = M \oplus N = M' \oplus N'.$$

As we have noted above, from the decomposition  $L = M \oplus N$ , it follows that in the space  $L$ , there exists a basis  $e_1, \dots, e_{2m}$  comprising the bases of the subspaces  $M$  and  $N$  in which the matrix of the bilinear form  $\varphi$  is equal to (11.32). The second decomposition  $L = M' \oplus N'$  gives us a similar basis  $e'_1, \dots, e'_{2m}$ .

Let us define the transformation  $\mathcal{U}$  by the action on the vectors of the basis  $e_1, \dots, e_{2m}$  according to the formula  $\mathcal{U}(e_i) = e'_i$  for all  $i = 1, \dots, 2m$ . It is obvious that then the image  $\mathcal{U}(M)$  is equal to  $M'$ . Furthermore, for any two vectors  $x = x_1 e_1 + \dots + x_{2m} e_{2m}$  and  $y = y_1 e_1 + \dots + y_{2m} e_{2m}$ , their images  $\mathcal{U}(x)$  and  $\mathcal{U}(y)$  have, in the basis  $e'_1, \dots, e'_{2m}$ , decompositions with the same coordinates:  $\mathcal{U}(x) = x_1 e'_1 + \dots + x_{2m} e'_{2m}$  and  $\mathcal{U}(y) = y_1 e'_1 + \dots + y_{2m} e'_{2m}$ . From this it follows that

$$\varphi(\mathcal{U}(x), \mathcal{U}(y)) = \sum_{i=1}^{2m} x_i y_{m+i} = \varphi(x, y),$$

showing that  $\mathcal{U}$  is an orthogonal transformation.  $\square$

Let us note that Theorem 11.16 does not assert the uniqueness of such a transformation  $\mathcal{U}$ . In fact, such is not the case. Let us consider this question in more detail. Let  $\mathcal{U}_1$  and  $\mathcal{U}_2$  be the two orthogonal transformations that were the subject of Theorem 11.16. Applying to both sides of the equality  $\mathcal{U}_1(M) = \mathcal{U}_2(M)$  the transformation  $\mathcal{U}_1^{-1}$ , we obtain  $\mathcal{U}_0(M) = M$ , where  $\mathcal{U}_0 = \mathcal{U}_1^{-1}\mathcal{U}_2$  is also an orthogonal transformation. Our further considerations are based on the following result.

**Lemma 11.17** *Let  $M$  be an  $m$ -dimensional isotropic subspace of a  $2m$ -dimensional space  $L$ , and let  $\mathcal{U}_0 : L \rightarrow L$  be an orthogonal transformation taking  $M$  to itself. Then the transformation  $\mathcal{U}_0$  is proper.*

*Proof* By assumption,  $M$  is an invariant subspace of the transformation  $\mathcal{U}_0$ . This means that in an arbitrary basis of the space  $L$  whose first  $m$  vectors form a basis of  $M$ , the matrix of the transformation  $\mathcal{U}_0$  has the block form

$$U_0 = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}, \quad (11.36)$$

where  $A, B, C$  are square matrices of order  $m$ .

The orthogonality of the transformation  $\mathcal{U}_0$  is expressed by the relationship (11.35), in which, as we have seen, with the selection of a suitable basis, we may consider that relationship (11.32) is satisfied. Setting in (11.35) in place of  $U$  the matrix (11.36), we obtain

$$\begin{pmatrix} A^* & 0 \\ B^* & C^* \end{pmatrix} \cdot \begin{pmatrix} 0 & E \\ E & 0 \end{pmatrix} \cdot \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} = \begin{pmatrix} 0 & E \\ E & 0 \end{pmatrix}.$$

Multiplying the matrices on the left-hand side of this equality brings it into the form

$$\begin{pmatrix} 0 & A^*C \\ C^*A & D \end{pmatrix} = \begin{pmatrix} 0 & E \\ E & 0 \end{pmatrix}, \quad \text{where } D = C^*B + B^*C.$$

From this, we obtain in particular  $A^*C = E$ , and this means that  $|A^*| \cdot |C| = 1$ . But in view of  $|A^*| = |A|$ , from (11.36) we have  $|U_0| = |A| \cdot |C| = 1$ , as asserted.  $\square$

From Lemma 11.17 we deduce the following important corollary.

**Theorem 11.18** *If  $M$  and  $M'$  are two  $m$ -dimensional isotropic subspaces of a  $2m$ -dimensional space  $L$ , then the orthogonal transformations  $\mathcal{U} : L \rightarrow L$  taking one of these subspaces into the other are either all proper or all improper.*

*Proof* Let  $\mathcal{U}_1$  and  $\mathcal{U}_2$  be two orthogonal transformations such that  $\mathcal{U}_i(M) = M'$ . It is clear that then  $\mathcal{U}_i^{-1}(M') = M$ . Setting  $\mathcal{U}_0 = \mathcal{U}_1^{-1}\mathcal{U}_2$ , from the equality  $\mathcal{U}_1(M) = \mathcal{U}_2(M)$  we obtain that  $\mathcal{U}_0(M) = M$ . By Lemma 11.17,  $|\mathcal{U}_0| = 1$ , and from the relationship  $\mathcal{U}_0 = \mathcal{U}_1^{-1}\mathcal{U}_2$ , it follows that  $|\mathcal{U}_1| = |\mathcal{U}_2|$ .  $\square$



Theorem 11.18 determines in an obvious way a partition of the set of all  $m$ -dimensional isotropic subspaces  $M$  of a  $2m$ -dimensional space  $L$  into two families  $\mathfrak{M}_1$  and  $\mathfrak{M}_2$ . Namely,  $M$  and  $M'$  belong to one family if an orthogonal transformation  $\mathcal{U}$  taking one of these subspaces into the other (which always exists, by Theorem 11.16) is proper (it follows from Theorem 11.18 that this definition does not depend on the choice of a specific transformation  $\mathcal{U}$ ).

Now we can easily prove the following property, which was established in the previous section for  $m = 2$ , for any  $m$ .

**Theorem 11.19** *Two  $m$ -dimensional isotropic subspaces  $M$  and  $M'$  of a  $2m$ -dimensional space  $L$  belong to one family  $\mathfrak{M}_i$  if and only if the dimension of their intersection  $M \cap M'$  has the same parity as  $m$ .*

*Proof* Let us recall that natural numbers  $k$  and  $m$  have the same parity if  $k + m$  is even, or equivalently, if  $(-1)^{k+m} = 1$ . Recalling now the definition of the partition of the set of  $m$ -dimensional isotropic subspaces into families  $\mathfrak{M}_1$  and  $\mathfrak{M}_2$  and setting  $k = \dim(M \cap M')$ , we may formulate the assertion of the theorem as follows:

$$|\mathcal{U}| = (-1)^{k+m}, \quad (11.37)$$

where  $\mathcal{U}$  is an arbitrary orthogonal transformation taking  $M$  to  $M'$ , that is, a transformation such that  $\mathcal{U}(M) = M'$ .

Let us begin the proof of relationship (11.37) with the case  $k = 0$ , that is, the case that  $M \cap M' = \{0\}$ . Then in view of the equality  $\dim M + \dim M' = \dim L$ , the sum of subspaces  $M + M' = M \oplus M'$  coincides with the entire space  $L$ . This means that  $M'$  exhibits all the properties of the isotropic subspace  $N$  constructed for the proof of Theorem 11.12. In particular, there exist bases  $e_1, \dots, e_m$  in  $M$  and  $f_1, \dots, f_m$  in  $M'$  such that

$$\varphi(e_i, f_i) = 1 \quad \text{for } i = 1, \dots, m, \quad \varphi(e_i, f_j) = 0 \quad \text{for } i \neq j.$$

We shall determine the transformation  $\mathcal{U} : L \rightarrow L$  by the conditions  $\mathcal{U}(e_i) = f_i$  and  $\mathcal{U}(f_i) = e_i$  for all  $i = 1, \dots, m$ . It is clear that  $\mathcal{U}(M) = M'$  and  $\mathcal{U}(M') = M$ . It is equally easy to see that in the basis  $e_1, \dots, e_m, f_1, \dots, f_m$ , the matrices of the transformation  $\mathcal{U}$  and bilinear form  $\varphi$  coincide and have the form (11.32). Substituting the matrix (11.32) in place of  $U$  and  $\Phi$  into formula (11.35), we see that it is converted to a true equality, that is, the transformation  $\mathcal{U}$  is orthogonal.

On the other hand, we have, therefore, the equality  $|\mathcal{U}| = |\Phi| = (-1)^m$ . It is easy to convince oneself that  $|\Phi| = (-1)^m$  by transposing the rows of the matrix (11.32) with indices  $i$  and  $m + i$  for all  $i = 1, \dots, m$ . Here we shall carry out  $m$  transpositions and obtain the identity matrix of order  $2m$  with determinant 1. As a result, we arrive at the equality  $|\mathcal{U}| = (-1)^m$ , that is, at relationship (11.37) for  $k = 0$ .

Now let us examine the case  $k > 0$ . Let us define the subspace  $M_1 = M \cap M'$ . Then  $k = \dim M_1$ . By Theorem 11.12, there exists an  $m$ -dimensional isotropic subspace  $N \subset L$  such that  $L = M \oplus N$ . Let us choose in the subspace  $M$  a basis  $e_1, \dots, e_m$

such that its first  $k$  vectors  $\mathbf{e}_1, \dots, \mathbf{e}_k$  form a basis in  $M_1$ . Then clearly, we have the decomposition

$$M = M_1 \oplus M_2, \quad \text{where } M_1 = \langle \mathbf{e}_1, \dots, \mathbf{e}_k \rangle, M_2 = \langle \mathbf{e}_{k+1}, \dots, \mathbf{e}_m \rangle.$$

Above (see Lemma 11.13), we constructed the isomorphism  $\mathcal{F} : N \xrightarrow{\sim} M^*$  and with its help, defined a basis  $\mathbf{g}_1, \dots, \mathbf{g}_m$  in the space  $N$  by formula  $\mathcal{F}(\mathbf{g}_i) = \mathbf{f}_i$ , where  $\mathbf{f}_1, \dots, \mathbf{f}_m$  is a basis of the space  $M^*$ , the dual basis to  $\mathbf{e}_1, \dots, \mathbf{e}_m$ . We obviously have the decomposition

$$N = N_1 \oplus N_2, \quad \text{where } N_1 = \langle \mathbf{g}_1, \dots, \mathbf{g}_k \rangle, N_2 = \langle \mathbf{g}_{k+1}, \dots, \mathbf{g}_m \rangle,$$

where by our construction,  $\mathcal{F} : N_1 \xrightarrow{\sim} M_1^*$  and  $\mathcal{F} : N_2 \xrightarrow{\sim} M_2^*$ .

Let us consider the linear transformation  $\mathcal{U}_0 : L \rightarrow L$  defined by the formula

$$\begin{aligned} \mathcal{U}_0(\mathbf{e}_i) &= \mathbf{g}_i, & \mathcal{U}_0(\mathbf{g}_i) &= \mathbf{e}_i \quad \text{for } i = 1, \dots, k, \\ \mathcal{U}_0(\mathbf{e}_i) &= \mathbf{e}_i, & \mathcal{U}_0(\mathbf{g}_i) &= \mathbf{g}_i \quad \text{for } i = k+1, \dots, m. \end{aligned}$$

It is obvious that the transformation  $\mathcal{U}_0$  is orthogonal, and also  $\mathcal{U}_0^2 = \mathcal{E}$  and

$$\begin{aligned} \mathcal{U}_0(M_1) &= N_1, & \mathcal{U}_0(M_2) &= M_2, \\ \mathcal{U}_0(N_1) &= M_1, & \mathcal{U}_0(N_2) &= N_2. \end{aligned} \tag{11.38}$$

In the basis  $\mathbf{e}_1, \dots, \mathbf{e}_m, \mathbf{g}_1, \dots, \mathbf{g}_m$  that we constructed in the space  $L$ , the matrix of the transformation  $\mathcal{U}_0$  has the block form

$$U_0 = \begin{pmatrix} 0 & 0 & E_k & 0 \\ 0 & E_{m-k} & 0 & 0 \\ E_k & 0 & 0 & 0 \\ 0 & 0 & 0 & E_{m-k} \end{pmatrix},$$

where  $E_k$  and  $E_{m-k}$  are the identity matrices of orders  $k$  and  $m-k$ . As is evident,  $U_0$  becomes the identity matrix after the transposition of its rows with indices  $i$  and  $m+i$ ,  $i = 1, \dots, k$ . Therefore,  $|\mathcal{U}_0| = (-1)^k$ .

Let us prove that  $\mathcal{U}_0(M') \cap M = (\mathbf{0})$ . Since  $\mathcal{U}_0^2 = \mathcal{E}$ , this is equivalent to  $M' \cap \mathcal{U}_0(M) = (\mathbf{0})$ . Let us assume that  $\mathbf{x} \in M' \cap \mathcal{U}_0(M)$ . From the membership  $\mathbf{x} \in \mathcal{U}_0(M)$  and decomposition  $M = M_1 \oplus M_2$ , taking into account (11.38), it follows that  $\mathbf{x} \in N_1 \oplus M_2$ , that is,

$$\mathbf{x} = \mathbf{z}_1 + \mathbf{y}_2, \quad \text{where } \mathbf{z}_1 \in N_1, \mathbf{y}_2 \in M_2. \tag{11.39}$$

Thus for every vector  $\mathbf{y}_1 \in M_1$ , we have the equality

$$\varphi(\mathbf{x}, \mathbf{y}_1) = \varphi(\mathbf{z}_1, \mathbf{y}_1) + \varphi(\mathbf{y}_2, \mathbf{y}_1). \tag{11.40}$$

The left-hand side of equality (11.40) equals zero, since  $\mathbf{x} \in M'$ ,  $\mathbf{y}_1 \in M_1 \subset M'$ , and the subspace  $M'$  is isotropic with respect to  $\varphi$ . The second term  $\varphi(\mathbf{y}_2, \mathbf{y}_1)$

on the right-hand side is equal to zero, since  $y_i \in M_i \subset M$ ,  $i = 1, 2$ , and the subspace  $M$  is isotropic with respect to  $\varphi$ . Thus from relationship (11.40), it follows that  $\varphi(z_1, y_1) = 0$  for every vector  $y_1 \in M_1$ .

This last conclusion means that for the isomorphism  $\mathcal{F} : N_1 \xrightarrow{\sim} M_1^*$ , there corresponds to the vector  $z_1 \in N_1$ , a linear function on  $M_1$  that is identically equal to zero. But that can be the case only if the vector  $z_1$  itself is equal to  $\mathbf{0}$ . Thus in the decomposition (11.39), we have  $z_1 = \mathbf{0}$ , and therefore, the vector  $x = y_2$  is contained in the subspace  $M_2$ . On the other hand, by virtue of the inclusions  $M_2 \subset M$  and  $x \in M' \cap \mathcal{U}_0(M)$ , taking into account the definition of the subspace  $M_1 = M \cap M'$ , this vector is also contained in  $M_1$ . As a result, we obtain that  $x \in M_1 \cap M_2$ , while by virtue of the decomposition  $M = M_1 \oplus M_2$ , this means that  $x = \mathbf{0}$ .

Thus the subspaces  $\mathcal{U}_0(M')$  and  $M$  are included in the case  $k = 0$  already considered, and relationship (11.37) has been proved for them. By Theorem 11.16, there exists an orthogonal transformation  $\mathcal{U}_1 : L \rightarrow L$  such that  $\mathcal{U}_1(\mathcal{U}_0(M')) = M$ . Then, as we have proved,  $|\mathcal{U}_1| = (-1)^m$ . The orthogonal transformation  $\mathcal{U} = \mathcal{U}_1 \mathcal{U}_0$  takes the isotropic subspace  $M'$  to  $M$ , and for it we have the relationship

$$|\mathcal{U}| = |\mathcal{U}_1| \cdot |\mathcal{U}_0| = (-1)^m (-1)^k = (-1)^{k+m},$$

which completes the proof of the theorem.  $\square$

We note two corollaries to Theorem 11.19.

**Corollary 11.20** *The families  $\mathfrak{M}_1$  and  $\mathfrak{M}_2$  do not have an  $m$ -dimensional isotropic subspace in common.*

*Proof* Let us assume that two such  $m$ -dimensional isotropic subspaces  $M_1 \in \mathfrak{M}_1$  and  $M_2 \in \mathfrak{M}_2$  are to be found such that  $M_1 = M_2$ . Then we clearly have the equality  $\dim(M_1 \cap M_2) = m$ , and by Theorem 11.19,  $M_1$  and  $M_2$  cannot belong to different families  $\mathfrak{M}_1$  and  $\mathfrak{M}_2$ .  $\square$

**Corollary 11.21** *If two  $m$ -dimensional isotropic subspaces intersect in a subspace of dimension  $m - 1$ , then they belong to different families  $\mathfrak{M}_1$  and  $\mathfrak{M}_2$ .*

This follows from the fact that  $m$  and  $m - 1$  have opposite parity.

*Case 2.* Now we may proceed to an examination of the second case, in which the dimension of the space  $L$  is odd. It is considerably easier and can be reduced to the already considered case of even dimensionality.

In order to retain the previous notation used in the even-dimensional case, let us denote by  $\bar{L}$  the space of odd dimension  $2m + 1$  under consideration and let us embed it as a hyperplane in a space  $L$  of dimension  $2m + 2$ . Let us denote by  $F$  a nonsingular quadratic form on  $L$  and by  $\bar{F}$  its restriction to  $\bar{L}$ . Our further reasoning will be based on the following fact.

**Lemma 11.22** *For every nonsingular quadratic form  $F$  there exists a hyperplane  $\bar{L} \subset L$  such that the quadratic form  $\bar{F}$  is nonsingular.*

*Proof* In a complex projective space, all nonsingular quadratic forms are equivalent. And therefore, it suffices to prove the required assertion for any one form  $F$ . For  $F$ , let us take the nonsingular form (11.33) that we encountered previously with  $m$  replaced by  $m + 1$ . Thus for a vector  $\mathbf{x} \in L$  with coordinates  $(x_1, \dots, x_{2m+2})$ , we have

$$F(\mathbf{x}) = \sum_{i=1}^{m+1} x_i x_{m+1+i}. \quad (11.41)$$

Let us define a hyperplane  $\bar{L} \subset L$  by the equation  $x_1 = x_{m+2}$ . The coordinates in  $\bar{L}$  are collections  $(x_1, \dots, x_{m+1}, \check{x}_{m+2}, x_{m+3}, \dots, x_{2m+2})$ , where the symbol  $\check{\phantom{x}}$  indicates the omission of the coordinate underneath it, and the quadratic form  $\bar{F}$  in these coordinates takes the form

$$\bar{F}(\mathbf{x}) = x_1^2 + \sum_{i=2}^{m+1} x_i x_{m+1+i}. \quad (11.42)$$

The matrix of the quadratic form (11.42) has the block form

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \Phi & \\ 0 & & & \end{pmatrix},$$

where  $\Phi$  is the matrix from formula (11.32). Since the determinant  $|\Phi|$  is nonzero, it follows that the quadratic form (11.42) is nonsingular.  $\square$

We shall further investigate the  $m$ -dimensional subspaces  $\bar{M} \subset \bar{L}$ , isotropic with respect to the nonsingular quadratic form  $\bar{F}$ , which is the restriction to the hyperplane  $\bar{L}$  of the nonsingular quadratic form  $F$  given in the surrounding space  $L$ . Since in the complex projective space  $\bar{L}$  all nonsingular quadratic forms are equivalent, it follows that all our results will be valid for an arbitrary nonsingular quadratic form on  $\bar{L}$ .

Let us consider an arbitrary  $(m + 1)$ -dimensional subspace  $M \subset L$ , isotropic with respect to  $F$ , and let us set  $\bar{M} = M \cap \bar{L}$ . It is obvious that the subspace  $\bar{M} \subset \bar{L}$  is isotropic with respect to  $\bar{F}$ . Since in the space  $L$ , the hyperplane  $\bar{L}$  is defined by a single linear equation, it follows that either  $M \subset \bar{L}$  (and then  $\bar{M} = M$ ), or  $\dim \bar{M} = \dim M - 1 = m$ . But the first case is impossible, since  $\dim \bar{M} \leq \frac{1}{2} \dim \bar{L} = \frac{1}{2}(2m + 1)$ , and  $\dim M = m + 1$ . Thus there remains the second case:  $\dim \bar{M} = m$ . Let us show that such an association with an  $(m + 1)$ -dimensional isotropic subspace  $M \subset L$  of an  $m$ -dimensional isotropic subspace  $\bar{M} \subset \bar{L}$  gives all the subspaces  $\bar{M}$  of interest to us and in a certain sense, it is unique.

**Theorem 11.23** *For every  $m$ -dimensional subspace  $\overline{M} \subset \overline{L}$  isotropic with respect to  $\overline{F}$ , there exists an  $(m+1)$ -dimensional subspace  $M \subset L$ , isotropic with respect to  $F$ , such that  $\overline{M} = M \cap \overline{L}$ . Moreover, in each of the families  $\mathfrak{M}_1$  and  $\mathfrak{M}_2$  of subspaces isotropic with respect to  $F$ , there exists such an  $M$ , and it is unique.*

*Proof* Let us consider an arbitrary  $m$ -dimensional subspace  $\overline{M} \subset \overline{L}$ , isotropic with respect to  $\overline{F}$ , and let us denote by  $\overline{M}^\perp$  its orthogonal complement with respect to the symmetric bilinear form  $\varphi$  associated with the quadratic form  $F$  in the surrounding space  $L$ . According to our previous notation, it should have been denoted by  $\overline{M}_\varphi^\perp$ , but we shall suppress the subscript, since the bilinear form  $\varphi$  will be always one and the same. From relationship (7.75), which is valid for a nondegenerate (with respect to the form  $\varphi$ ) space  $L$  and an arbitrary subspace of it (p. 267), it follows that

$$\dim \overline{M}^\perp = \dim L - \dim \overline{M} = 2m + 2 - m = m + 2.$$

Let us denote by  $\tilde{\varphi}$  the restriction of the bilinear form  $\varphi$  to  $\overline{M}^\perp$ , and by  $\tilde{F}$  the restriction of the quadratic form  $F$  to  $\overline{M}^\perp$ . The forms  $\tilde{\varphi}$  and  $\tilde{F}$  are singular in general. By definition (p. 198), the radical of the bilinear form  $\tilde{\varphi}$  is equal to  $\overline{M}^\perp \cap (\overline{M}^\perp)^\perp = \overline{M}^\perp \cap \overline{M}$ . But since  $\overline{M}$  is isotropic, it follows that  $\overline{M} \subset \overline{M}^\perp$ , and therefore, the radical of the bilinear form  $\tilde{\varphi}$  coincides with  $\overline{M}$ . By relationship (6.17) from Sect. 6.2, the rank of the bilinear form  $\tilde{\varphi}$  is equal to

$$\dim \overline{M}^\perp - \dim (\overline{M}^\perp)^\perp = \dim \overline{M}^\perp - \dim \overline{M} = (m+2) - m = 2,$$

and in the subspace  $\overline{M}^\perp$ , we may choose a basis  $e_1, \dots, e_{m+2}$  such that its last  $m$  vectors are contained in  $\overline{M}$  (that is, in the radical of  $\tilde{\varphi}$ ), and the restriction of  $\varphi$  to  $\langle e_1, e_2 \rangle$  has matrix  $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ .

Thus we have the decomposition  $\overline{M}^\perp = \langle e_1, e_2 \rangle \oplus \overline{M}$ , where the restriction of the quadratic form  $F$  to  $\langle e_1, e_2 \rangle$  in our basis has the form  $x_1x_2$ , and the restriction of  $F$  to  $\overline{M}$  is identically equal to zero.

Let us set  $M_i = \overline{M} \oplus \langle e_i \rangle$ ,  $i = 1, 2$ . Then  $M_1$  and  $M_2$  are  $(m+1)$ -dimensional subspaces in  $L$ . It follows from this construction that the  $M_i$  are isotropic with respect to the bilinear form  $\varphi$ . Here  $M_i \cap \overline{L} = \overline{M}$ , since on the one hand, from considerations of dimensionality,  $M_i \not\subset \overline{L}$ , and on the other hand,  $\overline{M} \subset M_i$  and  $\overline{M} \subset \overline{L}$ . We have thus constructed two isotropic subspaces  $M_i \subset L$  such that  $M_i \cap \overline{L} = \overline{M}$ . That they belong to different families  $\mathfrak{M}_i$  and that in neither of these families are there any other subspaces with these properties, follows from Corollary 11.21.  $\square$

Thus we have shown that there exists a bijection between the set of  $m$ -dimensional isotropic subspaces  $\overline{M} \subset \overline{L}$  and each of the families  $\mathfrak{M}_i$  of  $(m+1)$ -dimensional isotropic subspaces  $M \subset L$ . This fact is expressed by saying that  $m$ -dimensional subspaces  $\overline{M} \subset \overline{L}$  isotropic with respect to a nonsingular quadratic form  $\overline{F}$  form a single family.

Of course, our partition of the set of isotropic subspaces into families is a matter of convention. It is mostly a tribute to tradition originating in the special cases considered in analytic geometry. However, it is possible to give a more precise meaning to this partition by describing these subspaces in terms of Plücker coordinates.

In the previous chapter, we showed that  $k$ -dimensional subspaces  $M$  of an  $n$ -dimensional space  $L$  are in one-to-one correspondence with the points of some projective algebraic variety  $G(k, n)$ , called the Grassmannian. Suppose we are given some nonsingular quadratic form  $F$  on the space  $L$ . Let us denote by  $I(k, n)$  the subset of points of the Grassmannian  $G(k, n)$  that correspond to the  $k$ -dimensional isotropic subspaces.

We shall state the following propositions without proof, since they relate not to linear algebra, but rather to algebraic geometry.<sup>4</sup>

**Proposition 11.24** *The set  $I(k, n)$  is a projective algebraic variety.*

In other words, this proposition asserts that the property of a subspace being isotropic can be described by certain homogeneous relationships among its Plücker coordinates.

A projective algebraic variety  $X$  is said to be *irreducible* if it cannot be represented in the form of a union  $X = X_1 \cup X_2$ , where  $X_i$  are projective algebraic varieties different from  $X$  itself.

Suppose the space  $L$  has odd dimension  $n = 2m + 1$ .

**Proposition 11.25** *The set  $I(m, 2m + 1)$  is an irreducible projective algebraic variety.*

Now let the space  $L$  have even dimension  $n = 2m$ . We shall denote by  $I_i(m, 2m)$  the subset of the projective algebraic variety  $I(m, 2m)$  whose points correspond to  $m$ -dimensional isotropic subspaces of the family  $\mathfrak{M}_i$ . Theorem 11.19 and its corollaries show that

$$I(m, 2m) = I_1(m, 2m) \cup I_2(m, 2m), \quad I_1(m, 2m) \cap I_2(m, 2m) = \emptyset.$$

This suggests the idea that the projective algebraic variety  $I(m, 2m)$  is reducible.

**Proposition 11.26** *The sets  $I_i(m, 2m)$ ,  $i = 1, 2$ , are irreducible projective algebraic varieties.*

Finally, we have the following assertion, which relates to the isotropism of a subspace whose dimension is less than maximal.

**Proposition 11.27** *For all  $k < n/2$ , the projective algebraic variety  $I(k, n)$  is irreducible.*

---

<sup>4</sup>The reader can find them, for example, in the book *Methods of Algebraic Geometry*, by Hodge and Pedoe (Cambridge University Press, 1994).

## 11.4 Quadrics in a Real Projective Space

Let us consider a projective space  $\mathbb{P}(L)$ , where  $L$  is a real vector space. As before, we shall restrict our attention to the case of nonsingular quadrics. As we saw in Sect. 6.3 (formula (6.28)), a nonsingular quadratic form in a real space has the canonical form

$$x_0^2 + x_1^2 + \cdots + x_s^2 - x_{s+1}^2 - \cdots - x_n^2 = 0. \quad (11.43)$$

Here the index of inertia  $r = s + 1$  will be the same in every coordinate system in which the quadric is given by the canonical equation.

If we multiply equation (11.43) by  $-1$ , we obviously do not change the quadric that it defines, and therefore, we may assume that  $s + 1 \geq n - s$ , that is,  $s \geq (n - 1)/2$ . Moreover,  $s \leq n$ , but in the case  $s = n$ , from equation (11.43) we obtain  $x_0 = 0, x_1 = 0, \dots, x_n = 0$ , and there is no such point in projective space.

Thus, in contrast to a complex projective space, in a real projective space of given dimension  $n$ , there exists (up to a projective transformation) not one, but several nonsingular quadrics. However, there is only a finite number of them; they correspond to various values  $s$ , where we may assume that

$$\frac{n-1}{2} \leq s \leq n-1. \quad (11.44)$$

To be sure, it is still necessary to prove that the quadrics corresponding to the various values of  $s$  are not projectively equivalent. But we shall consider this question (in an even more complex situation) in the next section.

Thus the number of projectively inequivalent nonsingular quadrics in a real projective space of dimension  $n$  is equal to the number of integers  $s$  satisfying inequality (11.44). If  $n$  is odd,  $n = 2m + 1$ , then inequality (11.44) gives  $m \leq s \leq 2m$ , and the number of projectively inequivalent quadrics is equal to  $m + 1$ . And if  $n$  is even,  $n = 2m$ , then there are  $m$  of them. In particular, for  $n = 2$ , all nonsingular quadrics in the projective plane are projectively equivalent. The most typical example is the circle  $x^2 + y^2 = 1$ , which is contained entirely in the affine part of  $x_2 \neq 0$  if the equation is written as  $x_0^2 + x_1^2 - x_2^2 = 0$  in homogeneous coordinates  $(x_0 : x_1 : x_2)$  (here inhomogeneous coordinates are expressed by the formulas  $x = x_0/x_2, y = x_1/x_2$ ).

In three-dimensional projective space, there exist two types of projectively inequivalent quadrics. In homogeneous coordinates  $(x_0 : x_1 : x_2 : x_3)$ , one of them is given by the equation  $x_0^2 + x_1^2 + x_2^2 - x_3^2 = 0$ . Here we always have  $x_3 \neq 0$ , the quadric lies in the affine part, and it is given in inhomogeneous coordinates  $(x, y, z)$  by the equation  $x^2 + y^2 + z^2 = 1$ , where  $x = x_0/x_3, y = x_1/x_3, z = x_2/x_3$ . This quadric is a sphere. The second type is given by the equation  $x_0^2 + x_1^2 - x_2^2 - x_3^2 = 0$ . This is a hyperboloid of one sheet.

Their projective inequivalence can be seen at the very least from the fact that not a single real line lies on the first of them (the sphere), while on the second (hyperboloid of one sheet), there are two families each consisting of an infinite number of lines, called the rectilinear generatrices.

Of course, we can embed a real space  $L$  into a complex space  $L^{\mathbb{C}}$ , and similarly, embed  $\mathbb{P}(L)$  into  $\mathbb{P}(L^{\mathbb{C}})$ . Therefore, everything that was said in Sect. 11.3 about

isotropic subspaces is applicable in our case. However, although our quadric is real, the isotropic subspaces obtained in this way can turn out to be complex. The single exception is the case in which if the number  $n$  is odd, then  $s = (n - 1)/2$ , or if  $n$  is even, then  $s = n/2$ .

In the first instance, we may combine the coordinates into pairs  $(x_i, x_{s+1+i})$  and set  $u_i = x_i + x_{s+1+i}$  and  $v_i = x_i - x_{s+1+i}$ . Then taking into account the equalities

$$x_i^2 - x_{s+1+i}^2 = (x_i + x_{s+1+i})(x_i - x_{s+1+i}),$$

equation (11.43) can be written in the form

$$u_0 v_0 + u_1 v_1 + \cdots + u_s v_s = 0. \quad (11.45)$$

But this is the case of the quadric (11.33), which we considered in the previous section. It is easy to see that the reasoning used in Sect. 11.3 gives us a description of the real subspaces of a quadric.

The case  $s = n/2$  for even  $n$  also does not remove us from the realm of real subspaces and also leads to the case considered in the previous section. Moreover, if the equation of a quadric has the form (11.45) over an arbitrary field  $\mathbb{K}$  of characteristic different from 2, then the reasoning from the previous section remains in force.

In the general case, it is still possible to determine the dimensions of the spaces contained in a quadric. For this, we may make use of considerations already used in the proof of the law of inertia (Theorem 6.17 from Sect. 6.3). There we observed that the index of inertia (in the given case, the index of inertia of the quadratic form from (11.43), equal to  $s + 1$ ) coincides with the maximal dimension of the subspaces  $L'$  on which the restriction of the form is positive definite. (Let us note that this condition gives a geometric characteristic of the index of inertia, that is, it depends only on the set of solutions of the equation  $F(x) = 0$ , and not on the form  $F$  that defines it.)

Indeed, let the quadric  $Q$  be given by the equation  $F(x) = 0$ . If the restriction  $F'$  of the form  $F$  to the subspace  $L'$  is positive definite, then it is clear that  $Q \cap \mathbb{P}(L') = \emptyset$ . Thus if we are dealing with a projective space  $\mathbb{P}(L)$ , where  $\dim L = n + 1$ , then in  $L$  there exists a subspace  $\bar{L}$  of dimension  $s + 1$  such that the restriction of the form  $F$  to it is positive definite. This means that  $Q \cap \mathbb{P}(\bar{L}) = \emptyset$  (however, such a subspace  $\bar{L}$  is also easily determined explicitly on the basis of equation (11.43)). If  $L' \subset L$  is a subspace such that  $\mathbb{P}(L') \subset Q$ , then  $L' \cap \bar{L} = \{0\}$ . Hence by Corollary 3.42, we obtain the inequality  $\dim \bar{L} + \dim L' \leq \dim L = n + 1$ . Consequently,  $\dim L' + s + 1 \leq n + 1$ , and this means that  $\dim L' \leq n - s$ . Thus for the space  $\mathbb{P}(L')$  belonging to the quadric given by equation (11.43), we obtain  $\dim L' \leq n - s$  and therefore  $\dim \mathbb{P}(L') \leq n - s - 1$ .

On the other hand, it is easy to produce a subspace of dimension  $n - s - 1$  actually belonging to the quadric (11.43). To this end, let us combine in pairs the unknowns appearing in equation (11.43) with different signs and let us equate the unknowns in one pair, for example  $x_0 = x_{s+1}$ , and so on. Since we have assumed that  $s + 1 \geq n - s$ , we may form  $n - s$  such pairs, and therefore, we obtain  $n - s$  linear equations. How many unknowns remain? Since we have combined  $2(n - s)$  unknowns into



pairs, and in all there were  $n + 1$  of them, there remain  $n + 1 - 2(n - s)$  unknowns (it is possible that this number will be equal to zero). Thus we obtain

$$(n - s) + n + 1 - 2(n - s) = n + 1 - (n - s)$$

linear equations in coordinates in the space  $L$ . Since different unknowns occur in all these equations, these equations are linearly independent and determine in  $L$  a subspace  $L'$  of dimension  $n - s$ . Then  $\dim \mathbb{P}(L') = n - s - 1$ . Of course, since  $L'$  is contained in  $Q$ , an arbitrary subspace  $\mathbb{P}(L'') \subset \mathbb{P}(L')$  for  $L'' \subset L'$  is also contained in  $Q$ . Thus in the quadric  $Q$  are contained subspaces of all dimensions  $r \leq n - s - 1$ .

We have therefore proved the following result.

**Theorem 11.28** *If a nonsingular quadric  $Q$  in a real projective space of dimension  $n$  is given by the equation  $F(x_0, \dots, x_n) = 0$  and the index of inertia of the quadratic form  $F$  is equal to  $s + 1$ , then in  $Q$  are contained projective subspaces only of dimension  $r \leq n - s - 1$ , and for each such number  $r$  there can be found in  $Q$  a projective subspace of dimension  $r$  (when  $s + 1 \geq n - r$ , which is always possible to attain without changing the quadric  $Q$ , but changing only the quadratic form  $F$  that determines it to  $-F$ ).*

We have already considered an example of a quadric in real three-dimensional projective space ( $n = 3$ ). Let us note that in this space there are only two nonempty quadrics: for  $s = 1$  and  $s = 2$ .

For  $s = 2$ , equation (11.43) can be written in the form

$$x_0^2 + x_1^2 + x_2^2 = x_3^2. \quad (11.46)$$

As we have already said, for points of a real quadric, we have  $x_3 \neq 0$ . This means that our quadric is entirely contained in this affine subset. Setting  $x = x_0/x_3$ ,  $y = x_1/x_3$ ,  $z = x_2/x_3$ , we shall write its equation in the form

$$x^2 + y^2 + z^2 = 1.$$

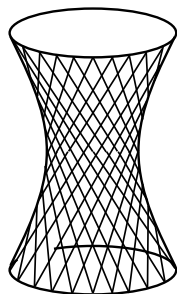
This is the familiar two-dimensional sphere  $S^2$  in three-dimensional Euclidean space. Let us discover what lines lie on it. Of course, no real line can lie on a sphere, since every line has points that are arbitrarily distant from the center of the sphere, while for all points of the sphere, their distance from the center of the sphere is equal to 1. Therefore, we can be talking only about complex lines of the space  $\mathbb{P}(L^{\mathbb{C}})$ . If in equation (11.46) we make the substitution  $x_2 = iy$ , where  $i$  is the imaginary unit, we obtain the equation  $x_0^2 + x_1^2 - y^2 - x_3^2 = 0$ , which in the new coordinates

$$u_0 = x_0 + y, \quad v_0 = x_0 - y, \quad u_1 = x_1 + x_3, \quad v_1 = x_1 - x_3$$

takes the form

$$u_0 v_0 + u_1 v_1 = 0. \quad (11.47)$$

**Fig. 11.3** Hyperboloid of one sheet



We studied such an equation in Sect. 11.2 (see Example 11.11). As an example of a line lying in the given quadric, we may take the line given by equations (11.25):  $u_0 = \lambda u_1$ ,  $v_0 = -\lambda^{-1} v_1$  with arbitrary complex number  $\lambda \neq 0$  and arbitrary  $u_1, v_1$ . In general, such a line contains *not a single real point* of our quadric (that is, points corresponding to real values of the coordinates  $x_0, \dots, x_3$ ). Indeed, if the number  $\lambda$  is not real, then the equality  $u_0 = \lambda u_1$  contradicts the fact that  $u_0$  and  $u_1$  are real. The case  $u_0 = u_1 = 0$  would correspond to a point with coordinates  $x_1 = x_3 = 0$ , for which  $x_0^2 + x_2^2 = 0$ , that is, all  $x_i$  are equal to zero.

Thus on the sphere lies a set of complex lines containing not a single real point. If desired, all of them could be described by formulas (11.27) and (11.28) after changes in coordinates that we described earlier. However, of greater interest are the complex lines lying on the sphere and containing at least one real point. For each such line  $l$  containing a real point of the sphere  $P$ , the *complex conjugate* line  $\bar{l}$  (that is, consisting of points  $\bar{Q}$ , where  $Q$  takes values on the line  $l$ ) also lies on the sphere and contains the point  $P$ . But by Theorem 11.19, through every point  $P$  pass exactly two lines (even if complex). We see that *through every point of the sphere there pass exactly two complex lines, which are the complex conjugates of each other*.

Finally, the case  $s = 1$  leads to the equation

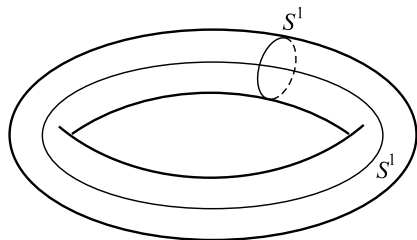
$$x_0^2 + x_1^2 - x_2^2 - x_3^2 = 0, \quad (11.48)$$

which after a change of coordinates

$$u_0 = x_0 + x_1, \quad v_0 = x_0 - x_1, \quad u_1 = x_2 + x_3, \quad v_1 = x_2 - x_3,$$

also assumes the form (11.47). For this equation, we have described all the lines contained in a quadric by formulas (11.27) and (11.28), where clearly, real values must be assigned to the parameters  $a, b, c, d$  in these formulas. In this case, the obtained quadric is a hyperboloid of one sheet, and the lines are its rectilinear generatrices. See Fig. 11.3.

Let us visualize what this surface looks like; that is, let us find a more familiar set that is homeomorphic to this surface. To this end, let us choose one line in each family of rectilinear generatrices: in the first,  $l_0$ ; in the second,  $l_1$ . As we saw in Sect. 9.4, every projective line is homeomorphic to the circle  $S^1$ . On the other hand,

**Fig. 11.4** A torus

every line in the second family of generatrices is uniquely determined by its point of intersection with the line  $l_0$ , and similarly, every line of the first family is determined by its point of intersection with the line  $l_1$ . Finally, through every point of the surface pass exactly two lines: one from the first family of generatrices, and the other from the second.

Thus is established a bijection between the points of a quadric given by equation (11.48) and pairs of points  $(\mathbf{x}, \mathbf{y})$ , where  $\mathbf{x} \in l_0$ ,  $\mathbf{y} \in l_1$ , that is, the set  $S^1 \times S^1$ . It is easily ascertained that this bijection is a homeomorphism. The set  $S^1 \times S^1$  is called a *torus*. It is most simply represented as the surface obtained by rotating a circle about an axis lying in the same plane as the circle but not intersecting it. See Fig. 11.4. Such a surface looks like the surface of a bagel. As a result, we obtain that the quadric given by equation (11.48) in three-dimensional real projective space is homeomorphic to a torus. See Fig. 11.4.

## 11.5 Quadrics in a Real Affine Space

Now we proceed to the study of quadrics in a real affine space  $(V, L)$ . Let us choose in this space a frame of reference  $(O; \mathbf{e}_1, \dots, \mathbf{e}_n)$ . Then every point  $A \in V$  is given by its coordinates  $(x_1, \dots, x_n)$ . A *quadric* is the set of all points  $A \in V$  such that

$$F(x_1, \dots, x_n) = 0, \quad (11.49)$$

where  $F$  is some second-degree polynomial. There is now no reason to consider the polynomial  $F$  to be homogeneous (as was the case in a projective space).

Collecting in  $F(\mathbf{x})$  terms of the second, first, and zeroth degrees, we shall write them in the form

$$F(\mathbf{x}) = \psi(\mathbf{x}) + f(\mathbf{x}) + c, \quad (11.50)$$

where  $\psi(\mathbf{x})$  is a quadratic form,  $f(\mathbf{x})$  is a linear form, and  $c$  is a scalar. The quadrics  $F(\mathbf{x}) = 0$  thus obtained for  $n = 2$  and  $3$  represent the curves and surfaces of order two studied in courses in analytic geometry.

Let us note that according to our definition of a quadric as a set of points satisfying relationship (11.49), we obtain even in the simplest cases,  $n = 2$  and  $3$ , sets that generally do not belong to curves or surfaces of degree two. The same “strange”

examples show that dissimilar-looking second-degree polynomials can define one and the same quadric, that is, the solution set of equation (11.49).

For example, in real three-dimensional space with coordinates  $x, y, z$ , the equation  $x^2 + y^2 + z^2 + c = 0$  has no solution in  $x, y, z$  if  $c > 0$ , and therefore for any  $c > 0$ , it defines the empty set. Another example is the equation  $x^2 + y^2 = 0$ , which is satisfied only with  $x = y = 0$  but for all  $z$ , that is, this equation defines a line, namely the  $z$ -axis. But the same line ( $z$ -axis) is defined, for example, by the equation  $ax^2 + by^2 = 0$  with any numbers  $a$  and  $b$  of the same sign.

Let us prove that if we exclude such “pathological” cases, then every quadric is defined by an equation that is unique up to a nonzero constant factor. Here it will be convenient to consider the empty set a special case of an affine subspace.

**Theorem 11.29** *If a quadric  $Q$  does not coincide with a set of points of any affine subspace and can be given by two different equations  $F_1(\mathbf{x}) = 0$  and  $F_2(\mathbf{x}) = 0$ , where the  $F_i$  are second-degree polynomials, then  $F_2 = \lambda F_1$ , where  $\lambda$  is some nonzero real number.*

*Proof* Since by the given condition, the quadric  $Q$  is not empty, it must contain some point  $A$ . By Theorem 8.14, there exists another point  $B \in Q$  such that the line  $l$  passing through  $A$  and  $B$  does not lie entirely in  $Q$ .

Let us select in the affine space  $V$ , a frame of reference  $(O; \mathbf{e}_1, \dots, \mathbf{e}_n)$  in which the point  $O$  is equal to  $A$  and the vector  $\mathbf{e}_1$  is equal to  $\overrightarrow{AB}$ . The line passing through the points  $A$  and  $B$  consists of points with coordinates  $(x_1, 0, \dots, 0)$  for all possible real values  $x_1$ . Let us write down the equation  $F_i(\mathbf{x}) = 0$ ,  $i = 1, 2$ , defining our quadric after arranging terms in order of the degree of  $x_1$ . As a result, we obtain the equations

$$F_i(x_1, \dots, x_n) = a_i x_1^2 + f_i(x_2, \dots, x_n)x_1 + \psi_i(x_2, \dots, x_n) = 0, \quad i = 1, 2,$$

where  $f_i(x_2, \dots, x_n)$  and  $\psi_i(x_2, \dots, x_n)$  are inhomogeneous polynomials of first and second degree in the variables  $x_2, \dots, x_n$ . After defining  $f_i(0, \dots, 0) = f_i(\overline{O})$  and  $\psi_i(0, \dots, 0) = \psi_i(\overline{O})$ , we may say that the relationship

$$a_i x_1^2 + f_i(\overline{O})x_1 + \psi_i(\overline{O}) = 0 \tag{11.51}$$

holds for  $x_1 = 0$  (point  $A$ ) and for  $x_1 = 1$  (point  $B$ ), but does not hold identically for all real values  $x_1$ . From this it follows that  $\psi_i(\overline{O}) = 0$  and  $a_i + f_i(\overline{O}) = 0$ . This means that  $a_i \neq 0$ , for otherwise, we would obtain that relationship (11.51) was satisfied for all  $x_1$ . By multiplying the polynomial  $F_i$  by  $a_i^{-1}$ , we may assume that  $a_i = 1$ .

Let us denote by  $\overline{\mathbf{x}}$  the projection of the vector  $\mathbf{x}$  onto the subspace  $\langle \mathbf{e}_2, \dots, \mathbf{e}_n \rangle$  parallel to the subspace  $\langle \mathbf{e}_1 \rangle$ , that is,  $\overline{\mathbf{x}} = (x_2, \dots, x_n)$ . Then we may say that the two equations

$$x_1^2 + f_1(\overline{\mathbf{x}})x_1 + \psi_1(\overline{\mathbf{x}}) = 0 \quad \text{and} \quad x_1^2 + f_2(\overline{\mathbf{x}})x_1 + \psi_2(\overline{\mathbf{x}}) = 0, \tag{11.52}$$

where  $f_i(\bar{\mathbf{x}})$  are first-degree polynomials and  $\psi_i(\bar{\mathbf{x}})$  are second-degree polynomials of the vector  $\bar{\mathbf{x}}$ , have identical solutions. Furthermore, we know that they both have two solutions,  $x_1 = 0$  and  $x_1 = 1$ , for  $\bar{\mathbf{x}} = \mathbf{0}$ , that is, the discriminant of each quadratic trinomial

$$p_i(x_1) = x_1^2 + f_i(\bar{\mathbf{x}})x_1 + \psi_i(\bar{\mathbf{x}}), \quad i = 1, 2,$$

with coefficients depending on the vector  $\bar{\mathbf{x}}$ , for  $\bar{\mathbf{x}} = \mathbf{0}$ , is positive.

The coefficients of the trinomial  $p_i(x_1)$  can be viewed as polynomials in the variables  $x_2, \dots, x_n$ , that is, the coordinates of the vector  $\bar{\mathbf{x}}$ . Consequently, the discriminant of the trinomial  $p_i(x_1)$  is also a polynomial in the variables  $x_2, \dots, x_n$ , and therefore, it depends on them continuously. From the definition of continuity, it follows that there exists a number  $\varepsilon > 0$  such that the discriminant of each trinomial  $p_i(x_1)$  is positive for all  $\bar{\mathbf{x}}$  such that  $|x_2| < \varepsilon, \dots, |x_n| < \varepsilon$ . This condition can be written compactly in the form of the single inequality  $|\bar{\mathbf{x}}| < \varepsilon$ , assuming that the space of vectors  $\bar{\mathbf{x}}$  is somehow converted into a Euclidean space in which is defined the length of a vector  $|\bar{\mathbf{x}}|$ . For example, it can be defined by the relationship  $|\bar{\mathbf{x}}|^2 = x_2^2 + \dots + x_n^2$ .

Thus the quadratic trinomials  $p_i(x_1)$  with leading coefficient 1 and coefficients  $f_i(\bar{\mathbf{x}})$  and  $\psi_i(\bar{\mathbf{x}})$ , depending continuously on  $\bar{\mathbf{x}}$ , each have two roots for all  $|\bar{\mathbf{x}}| < \varepsilon$ . But as is known from elementary algebra, such trinomials coincide. Therefore,  $f_1(\bar{\mathbf{x}}) = f_2(\bar{\mathbf{x}})$  and  $\psi_1(\bar{\mathbf{x}}) = \psi_2(\bar{\mathbf{x}})$  for all  $|\bar{\mathbf{x}}| < \varepsilon$ . Hence on the basis of the following lemma, we obtain that these equalities are satisfied not only for  $|\bar{\mathbf{x}}| < \varepsilon$ , but in general for all vectors  $\bar{\mathbf{x}}$ .  $\square$

**Lemma 11.30** *If for some number  $\varepsilon > 0$ , the polynomials  $f(\bar{\mathbf{x}})$  and  $g(\bar{\mathbf{x}})$  coincide for all  $\bar{\mathbf{x}}$  such that  $|\bar{\mathbf{x}}| < \varepsilon$ , then they coincide identically for all  $\bar{\mathbf{x}}$ .*

*Proof* Let us represent each of the polynomials  $f(\bar{\mathbf{x}})$  and  $g(\bar{\mathbf{x}})$  as a sum of homogeneous terms:

$$f(\bar{\mathbf{x}}) = \sum_{k=0}^N f_k(\bar{\mathbf{x}}), \quad g(\bar{\mathbf{x}}) = \sum_{k=0}^N g_k(\bar{\mathbf{x}}). \quad (11.53)$$

Let us set  $\bar{\mathbf{x}} = \alpha \bar{\mathbf{y}}$ , where  $|\bar{\mathbf{y}}| < \varepsilon$  and the number  $\alpha$  is in  $[0, 1]$ . Then the condition  $|\bar{\mathbf{x}}| < \varepsilon$  is clearly satisfied, and this means that  $f(\bar{\mathbf{x}}) = g(\bar{\mathbf{x}})$ . Setting  $\bar{\mathbf{x}} = \alpha \bar{\mathbf{y}}$  in equality (11.53), we obtain

$$\sum_{k=0}^N \alpha^k f_k(\bar{\mathbf{y}}) = \sum_{k=0}^N \alpha^k g_k(\bar{\mathbf{y}}). \quad (11.54)$$

On the one hand, equality (11.54) holds for all  $\alpha \in [0, 1]$ , of which there are infinitely many. On the other hand, (11.54) represents an equality between two polynomials in the variable  $\alpha$ . As is well known, polynomials of a single variable taking the same values for an infinite number of values of the variable coincide identically, that is, they have the same coefficients. Therefore, we obtain the equalities

$f_k(\bar{\mathbf{y}}) = g_k(\bar{\mathbf{y}})$  for all  $k = 0, \dots, N$  and all  $\bar{\mathbf{y}}$  for which  $|\bar{\mathbf{y}}| < \varepsilon$ . But since the polynomials  $f_k$  and  $g_k$  are homogeneous, it follows that these equalities hold in general for all  $\bar{\mathbf{y}}$ .

Indeed, every vector  $\bar{\mathbf{y}}$  can be represented in the form  $\bar{\mathbf{y}} = \alpha \bar{\mathbf{z}}$  with some scalar  $\alpha$  and vector  $\bar{\mathbf{z}}$  for which  $|\bar{\mathbf{z}}| < \varepsilon$ . For example, it suffices to set  $\alpha = (2/\varepsilon)|\bar{\mathbf{y}}|$ . Consequently, we obtain  $f_k(\bar{\mathbf{z}}) = g_k(\bar{\mathbf{z}})$ . But if we multiply both sides of this equality by  $\alpha^k$  and invoke the homogeneity of  $f_k$  and  $g_k$ , we obtain the equality  $f_k(\alpha \bar{\mathbf{z}}) = g_k(\alpha \bar{\mathbf{z}})$ , that is,  $f_k(\bar{\mathbf{y}}) = g_k(\bar{\mathbf{y}})$ , which is what was to be proved.  $\square$

Let us note that we might have posed this same question about the uniqueness of the correspondence between quadrics and their defining equations with regard to quadrics in projective space. But in projective space, the polynomial defining a quadric is homogeneous, and this question can be resolved even more easily. So that we wouldn't have to repeat ourselves, we have considered the question in the more complex situation.

Let us now investigate a question that is considered already in a course in analytic geometry for spaces of dimension 2 and 3: into what simplest form can equation (11.49) be brought by a suitable choice of frame of reference in an affine space of arbitrary dimension  $n$ ? This question is equivalent to the following: under what conditions can two quadrics be transformed into each other by a nonsingular affine transformation?

We shall consider quadrics in an affine space  $(V, \mathcal{L})$  of dimension  $n$ , assuming that for smaller values of  $n$ , this problem has already been solved. In this regard, we shall not consider quadrics that are *cylinders*, that is, having the form

$$Q = h^{-1}(Q'),$$

where  $(h, \mathcal{A})$  is an affine transformation of the space  $(V, \mathcal{L})$  into the affine space  $(V', \mathcal{L}')$  of dimension  $m < n$ , and  $Q'$  is some subset of  $V'$ . Let us ascertain that in this case,  $Q'$  is a quadric in  $V'$ .

Let the quadric  $Q$  in a coordinate system associated with some frame of reference of the affine space  $V$  be defined by the second-degree equation  $F(x_1, \dots, x_n) = 0$ . Let us choose in the  $m$ -dimensional affine space  $V'$  some frame of reference  $(O'; \mathbf{e}'_1, \dots, \mathbf{e}'_m)$ . Then  $\mathbf{e}'_1, \dots, \mathbf{e}'_m$  is a basis in the vector space  $\mathcal{L}'$ . In the definition of a cylinder, one has the condition  $\mathcal{A}(\mathcal{L}) = \mathcal{L}'$ . Let us denote by  $\mathbf{e}_1, \dots, \mathbf{e}_m$  vectors  $\mathbf{e}_i \in \mathcal{L}$  such that  $\mathcal{A}(\mathbf{e}_i) = \mathbf{e}'_i$ ,  $i = 1, \dots, m$ , and let us consider the subspace  $\mathcal{M} = \langle \mathbf{e}_1, \dots, \mathbf{e}_m \rangle$  that they span. By Corollary 3.31, there exists a subspace  $\mathcal{N} \subset \mathcal{L}$  such that  $\mathcal{L} = \mathcal{M} \oplus \mathcal{N}$ . Let  $O \in V$  be an arbitrary point such that  $h(O) = O'$ . Then in the coordinate system associated with the frame of reference  $(O'; \mathbf{e}'_1, \dots, \mathbf{e}'_m)$ , the projection of the space  $\mathcal{L}$  onto  $\mathcal{M}$  parallel to the subspace  $\mathcal{N}$  and the associated projection  $h$  of the affine space  $V$  onto  $V'$  are defined by the condition

$$h(x_1, \dots, x_n) = (x'_1, \dots, x'_m),$$

where  $x'_i$  are the coordinates of  $(O'; \mathbf{e}'_1, \dots, \mathbf{e}'_m)$ , the associated frame of reference. Then the fact that  $Q$  is a quadric means that its second-degree equation

$F(x_1, \dots, x_n) = 0$  is satisfied irrespective of the values that we have substituted for the variables  $x_{m+1}, \dots, x_n$  if the point with coordinates  $(x_1, \dots, x_m)$  belongs to the set  $Q'$ . For example, we may set  $x_{m+1} = 0, \dots, x_n = 0$ . Then the equation  $F(x'_1, \dots, x'_n, 0, \dots, 0) = 0$  will be precisely the equation of the quadric  $Q'$ .

The same reasoning shows that if a polynomial  $F$  depends on fewer than  $n$  unknowns, then the quadric  $Q$  defined by the equation  $F(x) = 0$  is a cylinder. Therefore, in the sequel we shall consider only quadrics that are not cylinders. Our goal will be the classification of these quadrics using nonsingular affine transformations. Two quadrics that can be mapped one into the other by such a transformation are said to be *affinely equivalent*.

First of all, let us consider the effect of a translation on the equation of a quadric. Let the equation of the quadric  $Q$  in coordinates associated with some frame of reference  $(O; e_1, \dots, e_n)$  have the form

$$F(x) = \psi(x) + f(x) + c = 0, \quad (11.55)$$

where  $\psi(x)$  is a quadratic form,  $f(x)$  is a linear form, and  $c$  is a number. If  $\mathcal{T}_a$  is a translation by the vector  $a \in L$ , then the quadric  $\mathcal{T}_a(Q)$  is given by the equation

$$\psi(x + a) + f(x + a) + c = 0.$$

Let us consider how the equation of a quadric is transformed under these conditions. Let  $\varphi(x, y)$  be the symmetric bilinear form associated with the quadratic form  $\psi(x)$ , that is,  $\psi(x) = \varphi(x, x)$ . Then

$$\begin{aligned} \psi(x + a) &= \varphi(x + a, x + a) = \varphi(x, x) + 2\varphi(x, a) + \varphi(a, a) \\ &= \psi(x) + 2\varphi(x, a) + \psi(a). \end{aligned}$$

As a result, we obtain that after a translation  $\mathcal{T}_a$ :

- (a) The quadratic part  $\psi(x)$  does not change.
- (b) The linear part  $f(x)$  is substituted by  $f(x) + 2\varphi(x, a)$ .
- (c) The constant term  $c$  is substituted by  $c + f(a) + \psi(a)$ .

Using statement (b), then with the aid of a translation  $\mathcal{T}_a$ , it is sometimes possible to eliminate the first-degree terms in the equation of a quadric. More precisely, this is possible if there exists a vector  $a \in L$  such that

$$f(x) = -2\varphi(x, a) \quad (11.56)$$

for an arbitrary  $x \in L$ . By Theorem 6.3, any bilinear form  $\varphi(x, y)$  can be represented in the form  $\varphi(x, y) = (x, \mathcal{A}(y))$  via some linear transformation  $\mathcal{A}: L \rightarrow L^*$ . Then condition (11.56) can be written in the form  $(x, f) = -2(x, \mathcal{A}(a))$  for all  $x \in L$ , that is, in the form  $f = -2\mathcal{A}(a) = \mathcal{A}(-2a)$ . This means that the condition (11.56) amounts to the linear function  $f \in L^*$  being contained in the image of the transformation  $\mathcal{A}$ .

First of all, let us investigate those quadrics for which condition (11.56) is satisfied. In this case, there exists a frame of reference of the affine space in which the quadric can be represented by the equation

$$F(\mathbf{x}) = \psi(\mathbf{x}) + c = 0. \quad (11.57)$$

This equation exhibits a remarkable symmetry: it is invariant under a change of the vector  $\mathbf{x}$  into  $-\mathbf{x}$ . Let us investigate this further.

**Definition 11.31** Let  $V$  be an affine space and  $A$  a point of  $V$ . A *central symmetry* with respect to a point  $A$  is a mapping  $V \rightarrow V$  that maps each point  $B \in V$  to the point  $B' \in V$  such that  $\overrightarrow{AB'} = -\overrightarrow{AB}$ .

It is obvious that by this condition, the point  $B'$ , and therefore the mapping, is uniquely determined. A trivial verification shows that this mapping is an affine transformation and its linear part is equal to  $-\mathcal{E}$ .

**Definition 11.32** A set  $Q \subset V$  is said to be *centrally symmetric* with respect to a point  $A \in V$  if it is invariant under a central symmetry with respect to the point  $A$ , which in this case is called the *center* of the set  $Q$ .

It follows from the definition that a point  $A$  on a quadric is a center if and only if the quadric is transformed into itself by the linear transformation  $-\mathcal{E}$ , that is,  $\mathbf{x} \mapsto -\mathbf{x}$ , where  $\mathbf{x} = \overrightarrow{AX}$  for every point  $X$  of this quadric.

**Theorem 11.33** *If a quadric does not coincide with an affine space, is not a cylinder, and has a center, then the center is unique.*

*Proof* Let  $A$  and  $B$  be two distinct centers of the quadric  $Q$ . This means, by definition, that for every point  $X \in Q$ , there exists a point  $X' \in Q$  such that

$$\overrightarrow{AX} = -\overrightarrow{AX'}, \quad (11.58)$$

and for every point  $Y \in Q$ , there exists a point  $Y' \in Q$  such that

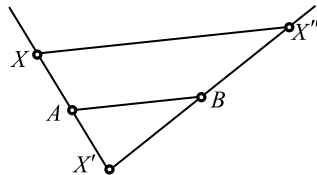
$$\overrightarrow{BY} = -\overrightarrow{BY'}. \quad (11.59)$$

Let us apply relationship (11.58) to an arbitrary point  $X \in Q$ , and relationship (11.59) to the associated point  $X' = Y$ . Let us denote the point  $Y'$  obtained as a result of these actions by  $X''$ . It is obvious that

$$\overrightarrow{XX''} = \overrightarrow{XA} + \overrightarrow{AB} + \overrightarrow{BX''}, \quad (11.60)$$

and from relationships (11.58) and (11.59), it follows that  $\overrightarrow{XA} = \overrightarrow{AX'}$  and  $\overrightarrow{BX''} = \overrightarrow{X'B}$ . Substituting the last expressions into (11.60), we obtain that  $\overrightarrow{XX''} = 2\overrightarrow{AB}$ . In other words, this means that if the vector  $\mathbf{e}$  is equal to  $2\overrightarrow{AB}$ , then the quadric  $Q$  is



**Fig. 11.5** Similar triangles

invariant under the translation  $\mathcal{T}_e$ ; see Fig. 11.5. This assertion also follows from an examination of the similar triangles  $ABX'$  and  $XX''X'$  in Fig. 11.5.

Since  $A \neq B$ , the vector  $e$  is nonnull. Let us choose an arbitrary frame of reference  $(O; e_1, \dots, e_n)$ , where  $e_1 = e$ . Let us set  $L' = \langle e_2, \dots, e_n \rangle$  and consider the affine space  $V' = (L', L')$  and mapping  $h: V \rightarrow V'$ , defined by the following conditions:  $h(O) = O$ ,  $h(A) = O$  if  $\overrightarrow{OA} = \lambda e$ , and  $h(A_i) = e_i$  if  $\overrightarrow{OA_i} = e_i$  ( $i = 2, \dots, n$ ). It is obvious that the mapping  $h$  is a projection and that the set  $Q$  is a cylinder. Since by our assumption, the quadric  $Q$  is not a cylinder, we have obtained a contradiction.  $\square$

Thus we obtain that by choosing a system of coordinates with the origin at the center of the quadric, one can define an arbitrary quadric satisfying the conditions of Theorem 11.33 by the equation

$$\psi(x_1, \dots, x_n) = c, \quad (11.61)$$

where  $\psi$  is a nonsingular quadratic form (in the case of a singular form  $\psi$ , the quadric would be a cylinder).

If  $c \neq 0$ , then we may assume that  $c = 1$  by multiplying both sides of equality (11.61) by  $c^{-1}$ . Finally, we may execute a linear transformation that preserves the origin and brings the quadratic form  $\psi$  into canonical form (6.22). As a result, the equation of the quadric takes the form

$$x_1^2 + \dots + x_r^2 - x_{r+1}^2 - \dots - x_n^2 = c, \quad (11.62)$$

where  $c = 0$  or 1, and the number  $r$  is the index of inertia of the quadratic form  $\psi$ .

If  $c = 0$  and  $r = 0$  or  $r = n$ , then it follows that  $x_1 = 0, \dots, x_n = 0$ , that is, the quadric consists of a single point, the origin, which contradicts the assumption made above that it does not coincide with some affine subspace. Likewise, for  $c = 1$  and  $r = 0$ , we obtain that  $-x_1^2 - \dots - x_n^2 = 1$ , and this is impossible for real  $x_1, \dots, x_n$ , so that the quadric consists of the empty set, which again contradicts our assumption.

We have thus proved the following assertion.

**Theorem 11.34** *If a quadric does not coincide with an affine subspace, is not a cylinder, and has a center, then in some coordinate system, it is defined by equation (11.62). Moreover,  $0 < r \leq n$ , and if  $c = 0$ , then  $r < n$ .*

In the case  $c = 0$ , it is possible, by multiplying the equation of a quadric by  $-1$ , to obtain that in (11.62), the number of positive terms is not less than the number of

negative terms, that is,  $r \geq n - r$ , or equivalently,  $r \geq n/2$ . In the sequel, we shall always assume that in the case  $c = 0$ , this condition is satisfied.

Theorem 11.34 asserts that every quadric that is not an affine subspace or a cylinder and that has a center can be transformed with the help of a suitable nonsingular affine transformation into a quadric given by equation (11.62). For  $c = 0$  (and only in this case), the quadric (11.62) is a *cone* (with its vertex at the origin), that is, for every one of its points  $\mathbf{x}$ , it also contains the entire line  $\langle \mathbf{x} \rangle$ . It is possible to indicate another characteristic property of a quadric given by equation (11.62) for  $c = 0$ : it is not smooth, while in the case  $c = 1$ , the quadric is smooth. This follows at once from the definition of singular points (the equalities  $F = 0$  and  $\frac{\partial F}{\partial x_i} = 0$ ).

Let us now consider quadrics without a center. Such a quadric  $Q$  is defined by the equation

$$F(\mathbf{x}) = \psi(\mathbf{x}) + f(\mathbf{x}) + c = 0, \quad (11.63)$$

where  $\psi(\mathbf{x})$  is a quadratic form,  $f(\mathbf{x})$  a linear form,  $c$  a scalar. As earlier, we shall write a symmetric bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$  corresponding to a quadratic form  $\psi(\mathbf{x})$  as  $\varphi(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathcal{A}(\mathbf{y}))$ , where  $\mathcal{A} : L \rightarrow L^*$  is a linear transformation. We have seen that for a quadric  $Q$  not to have a center is equivalent to the condition  $\mathbf{f} \notin \mathcal{A}(L)$ .

Let us choose an arbitrary basis  $\mathbf{e}_1, \dots, \mathbf{e}_{n-1}$  in the hyperplane  $L' = \langle \mathbf{f} \rangle^\perp$  defined in the space  $L$  by the linear homogeneous equation  $\mathbf{f}(\mathbf{x}) = 0$ , and let us extend this basis to a basis of the entire space  $L$  by means of a vector  $\mathbf{e}_n \perp L'$  such that  $\mathbf{f}(\mathbf{e}_n) = 1$  (here, of course, orthogonality is understood in the sense of being with respect to the bilinear form  $\varphi(\mathbf{x}, \mathbf{y})$ ). In the obtained frame of reference  $(O; \mathbf{e}_1, \dots, \mathbf{e}_n)$ , equation (11.63) can be written in the form

$$F(\mathbf{x}) = \psi'(x_1, \dots, x_{n-1}) + \alpha x_n^2 + x_n + c = 0, \quad (11.64)$$

where  $\psi'$  is the restriction of the quadratic form  $\psi$  to the hyperplane  $L'$ .

Let us now choose in  $L'$  a new basis  $\mathbf{e}'_1, \dots, \mathbf{e}'_{n-1}$ , in which the quadratic form  $\psi'$  has the canonical form

$$\psi'(x_1, \dots, x_{n-1}) = x_1^2 + \dots + x_r^2 - x_{r+1}^2 - \dots - x_{n-1}^2. \quad (11.65)$$

It is obvious that in this case, the coordinate origin  $O$  and the vector  $\mathbf{e}_n$  remain unchanged. If as a result, the quadratic form  $\psi'$  turned out to depend on fewer than  $n - 1$  variables, then the polynomial  $F$  in equation (11.63) would depend on fewer than  $n$  variables, and that, as we have seen, means that the quadric  $Q$  is a cylinder.

Let us show that in formula (11.64), the number  $\alpha$  is equal to 0. If  $\alpha \neq 0$ , then by virtue of the obvious relationship  $\alpha x_n^2 + x_n + c = \alpha(x_n + \beta)^2 + c'$ , where  $\beta = 1/(2\alpha)$  and  $c' = c - \beta/2$ , we obtain that via the translation  $\mathcal{T}_a$  by the vector  $\mathbf{a} = -\beta \mathbf{e}_n$ , equation (11.64) is transformed into

$$F(\mathbf{x}) = \psi'(x_1, \dots, x_{n-1}) + \alpha x_n^2 + c' = 0,$$

where  $\psi'$  has the form (11.65). But such an equation, as is easily seen, gives a quadric with a center.

Thus assuming that the quadric  $Q$  is not a cylinder and does not have a center, we obtain that its equation has the form

$$x_1^2 + \cdots + x_r^2 - x_{r+1}^2 - \cdots - x_{n-1}^2 + x_n + c = 0.$$

Now let us perform a translation  $\mathcal{T}_a$  by the vector  $a = -ce_n$ . As a result, the coordinates  $x_1, \dots, x_{n-1}$  are unchanged, while  $x_n$  is changed to  $x_n - c$ . In the new coordinates, the equation of the quadric assumes the form

$$x_1^2 + \cdots + x_r^2 - x_{r+1}^2 - \cdots - x_{n-1}^2 + x_n = 0. \quad (11.66)$$

By multiplying the equation of the quadric by  $-1$  and changing the coordinate  $x_n$  to  $-x_n$ , we can obtain that the number of positive squares in equation (11.66) is not less than the number of negative squares, that is,  $r \geq n - r - 1$ , or equivalently,  $r \geq (n - 1)/2$ .

We have thereby obtained the following result.

**Theorem 11.35** *Every quadric that is not an affine subspace or a cylinder and does not have a center can be given in some coordinate system by equation (11.66), where  $r$  is a number satisfying the condition  $(n - 1)/2 \leq r \leq n - 1$ .*

Thus by combining Theorems 11.34 and 11.35, we obtain the following result: *Every quadric that is not an affine subspace or a cylinder can be given in some coordinate system by equation (11.62) if it doesn't have a center and by equation (11.66) if it does have a center.* We call these equations *canonical*.

Theorems 11.34 and 11.35 do more than give the simplest form into which the equation of a quadric can be transformed through a suitable choice of coordinate system. Beyond that, it follows from these theorems that quadrics having a canonical form (11.62) or (11.66) can be affinely equivalent (that is, transformable into each other by a nonsingular affine transformation) only if their equations coincide.

On the way to proving this assertion, we shall first establish that quadrics defined by equation (11.66) never have a center. Indeed, writing the equation of a quadric in the form (11.50), we may say that it has a center only if  $f \in \mathcal{A}(L)$ . But a simple verification shows that this condition is not satisfied for quadrics defined by equation (11.66). Indeed, if in some basis  $e_1, \dots, e_n$  of the space  $L$ , the quadratic form  $\psi(x)$  is given as

$$x_1^2 + \cdots + x_r^2 - x_{r+1}^2 - \cdots - x_{n-1}^2,$$

then on choosing the dual basis  $f_1, \dots, f_n$ , of the dual space  $L^*$ , we obtain that the linear transformation  $\mathcal{A} : L \rightarrow L^*$  associated with  $\psi$  by the relationship  $\varphi(x, y) = (x, \mathcal{A}(y))$ , in which  $\varphi(x, y)$  is a symmetric bilinear form determined by the quadratic form  $\psi$ , has the form  $\mathcal{A}(e_i) = f_i$  for  $i = 1, \dots, r$ ,  $\mathcal{A}(e_i) = -f_i$  for  $i = r + 1, \dots, n - 1$ , and  $\mathcal{A}(e_n) = 0$ , and the linear form  $x_n$  coincides with  $f_n$ . Thus  $\mathcal{A}(L) = \langle f_1, \dots, f_{n-1} \rangle$  and  $f = f_n \notin \mathcal{A}(L)$ .

We may now formulate the fundamental theorem on the classification of quadrics with respect to nonsingular affine transformations.

**Theorem 11.36** *Any quadric that is not an affine subspace or cylinder can be represented in some coordinate system by the canonical equation (11.62) or (11.66), where the number  $r$  satisfies the conditions indicated in Theorems 11.34 and 11.35*

respectively. And conversely, every pair of quadrics having the canonical equation (11.62) or (11.66) in some coordinate systems can be transformed into each other by a nonsingular affine transformation only if their canonical equations coincide.

*Proof* Only the second part of the theorem remains to be proved. We have already seen that quadrics given by equations (11.62) and (11.66) cannot be mapped into each other by nonsingular affine transformations, since in the first case, the quadric has a center, while in the second case, it does not. Therefore, we may consider each case separately.

Let us begin with the first case. Let there be given two quadrics  $Q_1$  and  $Q_2$ , given by different canonical equations of the form (11.62) (we note that the canonical equations in this case differ by the value  $c = 0$  or  $1$  and index  $r$ ), and where  $Q_2 = g(Q_1)$ , with  $(g, \mathcal{A})$  a nonsingular affine transformation. By assumption, each quadric has a unique center, which in its chosen coordinate system coincides with the point  $O = (0, \dots, 0)$ .

Let us write down the transformation  $g$  in the form (8.19):  $g = \mathcal{T}_a g_0$ , where  $g_0(O) = O$ . By assumption,  $Q_2 = g(Q_1)$ , and this means that  $g(O) = O$ , that is, the vector  $a$  is equal to  $\mathbf{0}$ . In the equations of the quadrics, which we may write in the form  $F_i(\mathbf{x}) = \psi_i(\mathbf{x}) + c_i = 0$ ,  $i = 1$  and  $2$ , it is clear that  $F_i(\mathbf{0}) = c_i$ , and this means that the constants  $c_i$  coincide (in the sequel, we shall denote them by  $c$ ). Thus the equations of the quadrics  $Q_1$  and  $Q_2$  differ only in the quadratic part  $\psi_i(\mathbf{x})$ .

By Theorem 11.29, the transformation  $g$  takes the polynomial  $F_1(\mathbf{x}) - c$  into  $\lambda(F_2(\mathbf{x}) - c)$ , where  $\lambda$  is some nonzero real number. Consequently, the quadratic form  $\psi_1(\mathbf{x})$  is transformed into  $\lambda\psi_2(\mathbf{x})$  by the linear transformation  $\mathcal{A}$ . If we denote the indices of inertia of the quadratic forms  $\psi_i(\mathbf{x})$  by  $r_i$ , then from the law of inertia, it follows that either  $r_2 = r_1$  (for  $\lambda > 0$ ) or  $r_2 = n - r_1$  (for  $\lambda < 0$ ). In the case  $c = 0$ , we may assume that  $r_i \geq n/2$ , and the equality  $r_2 = n - r_1$  is possible only for  $r_2 = r_1$ . In the case  $c = 1$ , this same result follows from the fact that the transformation  $\mathcal{A}$  takes the polynomial  $\psi_1(\mathbf{x}) - 1$  into  $\lambda(\psi_1(\mathbf{x}) - 1)$ . Comparing the constant terms, we obtain  $\lambda = 1$ .

In the case that the quadric has no center, we may repeat the same arguments. We again obtain that the quadratic form  $\psi_1(\mathbf{x})$  is carried into  $\lambda\psi_2(\mathbf{x})$  by a nonsingular linear transformation. Since each form  $\psi_i(\mathbf{x})$  contains by assumption the term  $x_1^2$ , it follows that  $\lambda = 1$ , and from the law of inertia, it follows that  $r_2 = r_1$  (for  $\lambda > 0$ ), or  $r_2 = n - 1 - r_1$  (for  $\lambda < 0$ ). Since by assumption,  $r_i \geq (n - 1)/2$ , the equality  $r_2 = n - 1 - r_1$  is possible only for  $r_2 = r_1$ .  $\square$

Thus we see that in a real affine space of dimension  $n$ , there exists only a finite number of affinely inequivalent quadrics that are not affine subspaces or cylinders. Each of them is equivalent to a quadric that can be represented in the form of equation (11.62) or equation (11.66).

It is possible to compute the number of types of affinely inequivalent quadrics. Equation (11.62) for  $c = 1$  gives  $n$  possibilities. The remaining cases depend on the parity of the number  $n$ . If  $n = 2m$ , then equation (11.62) for  $c = 0$  gives  $m$  different types, and the same number is given by equation (11.66). Altogether, we obtain  $n + 2m = 2n$  different types in the case of even  $n$ . If  $n = 2m + 1$ , then equation

(11.62) for  $c = 0$  gives  $m$  different types, and the same number is given by equation (11.66). Altogether in this case we obtain  $n + 2m - 1 = 2n - 2$  different types. Thus in a real affine space of dimension  $n$ , the number of types of affinely inequivalent quadrics that are not affine subspaces or cylinders is equal to  $2n$  if  $n$  is even, and to  $2n - 2$  if  $n$  is odd.

*Remark 11.37* It is easy to see that the content of this section is reduced to the classification of second-degree polynomials  $F(x_1, \dots, x_n)$  up to a nonsingular affine transformation of the variables and multiplication by a nonzero scalar coefficient. The connection with the geometric object—the quadric—is established by Theorem 11.29. That we excluded from consideration the case of affine subspaces is related to the fact that we wished to emphasize the differences among the geometric objects that arise.

The assumption that the quadric was not a cylinder was made exclusively to emphasize the inductive nature of the classification. The limitations that we introduced could have been done without. By repeating precisely the same arguments, we obtain that an arbitrary set in  $n$ -dimensional affine space given by equating a second-degree polynomial in  $n$  variables—the coordinates of a point—to zero is affinely equivalent to one of the sets defined by the following equations:

$$x_1^2 + \dots + x_r^2 - x_{r+1}^2 - \dots - x_m^2 = 1, \quad 0 \leq r \leq m \leq n, \quad (11.67)$$

$$x_1^2 + \dots + x_r^2 - x_{r+1}^2 - \dots - x_m^2 = 0, \quad r \geq \frac{m}{2}, m \leq n, \quad (11.68)$$

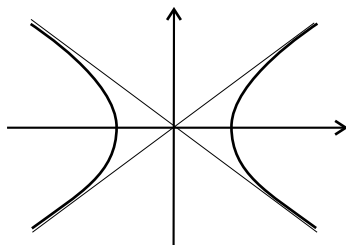
$$x_1^2 + \dots + x_r^2 - x_{r+1}^2 - \dots - x_{m-1}^2 + x_m = 0, \quad r \geq \frac{m-1}{2}, m < n. \quad (11.69)$$

After this, it is easy to see that in the case of (11.67) for  $r = 0$ , the empty set is obtained, while in the case (11.68) for  $r = 0$  or  $r = m$ , the result is an affine subspace. In the remaining cases, it is easy to find a line that intersects the given set in two distinct points and is not entirely contained in it. By virtue of Theorem 8.14, this means that such a set is not an affine subspace.

In conclusion, let us say a bit about the topological properties of affine quadrics.

If in equation (11.62), we have  $c = 1$  and the index of inertia  $r$  is equal to 1, then this equation can be rewritten in the form  $x_1^2 = 1 + x_2^2 + \dots + x_n^2$ , from which it follows that  $x_1^2 \geq 1$ , that is,  $x_1 \geq 1$  or  $x_1 \leq -1$ . Clearly, it is impossible for a point of the quadric whose coordinate  $x_1$  is greater than 1 to be continuously deformed into a point whose coordinate  $x_1$  is less than or equal to  $-1$  while remaining on the quadric (see the definition on p. xx). Therefore, a quadric in this case consists of two *components*, that is, it consists of two subsets such that no two points lying one in each of these subsets can be continuously deformed into each other while remaining on the quadric. It can be shown that each of these components is *path connected* (see the definition on p. xx), just as is every quadric given by equation (11.66).

The simplest example of a quadric consisting of two path-connected components is a hyperbola in the plane; see Fig. 11.6.

**Fig. 11.6** A hyperbola

The topological property that we described above has a generalization to quadrics defined by equation (11.62) for  $c = 1$  with smaller values of the index  $r$ , but still assuming that  $r \geq 1$ . Here we shall say a few words about them, without giving a rigorous formulation and also omitting proofs.

For  $r = 1$  we can find two points,  $(1, 0, \dots, 0)$  and  $(-1, 0, \dots, 0)$ , that cannot be transformed into each other by a continuous motion along the quadric (they could be given as the sphere  $x_1^2 = 1$  in one-dimensional space). For an arbitrary value of  $r$ , the quadric contains the sphere

$$x_1^2 + \dots + x_r^2 = 1, \quad x_{r+1} = 0, \quad \dots, \quad x_n = 0.$$

One can prove that this sphere cannot be contracted to a single point by continuous motion along the surface of the quadric. But for every  $m < r$  and continuous mapping  $f$  of the sphere  $S^{m-1} : y_1^2 + \dots + y_m^2 = 1$  into the quadric, the image of the sphere  $f(S^{m-1})$  can be contracted to a point by continuous motion along the quadric (it should be clear to the reader what is meant by continuous motion of a set along a quadric, something that we have already encountered in the case  $r = 1$ ).

## 11.6 Quadrics in an Affine Euclidean Space

It remains to us to consider nonsingular quadrics in an affine Euclidean space  $V$ . We shall, as before, exclude the cases in which the quadrics are affine subspaces or cylinders. The classification of such quadrics up to metric equivalence uses precisely the same arguments as those used in Sect. 11.5. To some extent, the results of that section can be applied in our case, since motions are affine transformations. Therefore, we shall only cursorily recall the line of reasoning.

Generalizing the statement of the problem, which goes back to analytic geometry (where cases  $\dim V = 2$  and  $3$  are considered), we shall say that two quadrics are *metrically equivalent* if they can be transformed into each other by some motion of the space  $V$ . This definition is a special case of metric equivalence of arbitrary metric spaces (see p. xxi), to which belong, as is easily verified, all quadrics in an affine Euclidean space.

First of all, let us consider quadrics given by equations whose linear part can be annihilated by a translation. These are quadrics that have a center (which, as we

have seen, is unique). Choosing a coordinate origin (that is, a point  $O$  of the frame of reference  $(O; \mathbf{e}_1, \dots, \mathbf{e}_n)$ ) in the center of the quadric, we bring its equation into the form

$$\psi(x_1, \dots, x_n) = c,$$

where  $\psi(x_1, \dots, x_n)$  is a nonsingular quadratic form,  $c$  a number. If  $c \neq 0$ , then by multiplying the equation by  $c^{-1}$ , we may assume that  $c = 1$ . For  $c = 0$ , the quadric is a cone.

Using an orthogonal transformation, the quadratic form  $\psi$  can be brought into canonical form

$$\psi(x_1, \dots, x_n) = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \dots + \lambda_n x_n^2,$$

where all the numbers  $\lambda_1, \dots, \lambda_n$  are nonzero, since by assumption, our quadric is nonsingular and is neither an affine subspace nor a cylinder, which means that the quadratic form  $\psi$  is nonsingular. Let us separate the positive numbers from the negative: suppose  $\lambda_1, \dots, \lambda_k > 0$  and  $\lambda_{k+1}, \dots, \lambda_n < 0$ . By tradition going back to analytic geometry, we shall set  $\lambda_i = a_i^{-2}$  for  $i = 1, \dots, k$  and  $\lambda_j = -a_j^{-2}$  for  $j = k+1, \dots, n$ , where all numbers  $a_1, \dots, a_n$  are positive.

Thus every quadric having a center is metrically equivalent to a quadric with equation

$$\left(\frac{x_1}{a_1}\right)^2 + \dots + \left(\frac{x_k}{a_k}\right)^2 - \left(\frac{x_{k+1}}{a_{k+1}}\right)^2 - \dots - \left(\frac{x_n}{a_n}\right)^2 = c, \quad (11.70)$$

where  $c = 0$  or  $1$ . For  $c = 0$ , multiplying equation (11.70) by  $-1$ , we may, as in the affine case, assume that  $k \geq n/2$ .

Now let us consider the case that the quadric

$$\psi(x_1, \dots, x_n) + f(x_1, \dots, x_n) + c = 0$$

does not have a center, that is,  $f \notin \mathcal{A}(L)$ , where  $\mathcal{A} : L \rightarrow L^*$  is the linear transformation associated with the quadratic form  $\psi$  by the relationship  $\varphi(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathcal{A}(\mathbf{y}))$ , in which  $\varphi(\mathbf{x}, \mathbf{y})$  is the symmetric bilinear form that gives the quadratic form  $\psi$ . In this case, it is easy to verify that as in Sect. 11.5, we can find an orthonormal basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of the space  $L$  such that

$$f(\mathbf{e}_1) = 0, \quad \dots, \quad f(\mathbf{e}_{n-1}) = 0, \quad f(\mathbf{e}_n) = 1,$$

and in the coordinate system determined by the frame of reference  $(O; \mathbf{e}_1, \dots, \mathbf{e}_n)$ , the quadric is given by the equation

$$\lambda_1 x_1^2 + \lambda_2 x_2^2 + \dots + \lambda_{n-1} x_{n-1}^2 + x_n + c = 0.$$

Through a translation by the vector  $-c\mathbf{e}_n$ , this equation can be brought into the form

$$\lambda_1 x_1^2 + \lambda_2 x_2^2 + \dots + \lambda_{n-1} x_{n-1}^2 + x_n = 0,$$

in which all the coefficients  $\lambda_i$  are nonzero, since the quadric is nonsingular and is not a cylinder.

If  $\lambda_1, \dots, \lambda_k > 0$  and  $\lambda_{k+1}, \dots, \lambda_{n-1} < 0$ , then by multiplying the equation of the quadric and the coordinate  $x_n$  by  $-1$  if necessary, we may assume that  $k \geq (n-1)/2$ . Setting, as previously,  $\lambda_i = a_i^{-2}$  for  $i = 1, \dots, k$  and  $\lambda_j = -a_j^{-2}$  for  $j = k+1, k+2, \dots, n-1$ , where  $a_1, \dots, a_n > 0$ , we bring the previous equation into the form

$$\left(\frac{x_1}{a_1}\right)^2 + \dots + \left(\frac{x_k}{a_k}\right)^2 - \left(\frac{x_{k+1}}{a_{k+1}}\right)^2 - \dots - \left(\frac{x_{n-1}}{a_{n-1}}\right)^2 + x_n = 0. \quad (11.71)$$

Thus every quadric in an affine Euclidean space is metrically equivalent to a quadric given by equation (11.70) (type I) or (11.71) (type II). Let us verify (under the given conditions and restriction on  $r$ ) that two quadrics of the form (11.70) or of the form (11.71) are metrically equivalent only if all the numbers  $a_1, \dots, a_n$  (for type I) and  $a_1, \dots, a_{n-1}$  (for type II) in their equations are the same. Here we may consider separately quadrics of type I and of type II, since they differ even from the viewpoint of affine equivalence.

By Theorem 8.39, every motion of an affine Euclidean space is the composition of a translation and an orthogonal transformation. As we saw in Sect. 11.5, a translation does not alter the quadratic part of the equation of a quadric. By Theorem 11.29, two quadrics are affinely equivalent only if the polynomials appearing in their equations differ by a constant factor. But for quadrics of type I for  $c = 1$ , this factor must be equal to 1. In the case of a quadric of type I for  $c = 0$ , multiplication by  $\mu > 0$  means that all the numbers  $a_i$  are multiplied by  $\mu^{-1/2}$ . For a quadric of type II, this factor must also be equal to 1 in order to preserve the coefficient 1 in the linear term  $x_n$ .

Thus we see that if we exclude quadrics of type I with constant term  $c = 0$  (a cone), then the quadratic parts of the equations must be quadratic forms equivalent with respect to orthogonal transformations. But the numbers  $\lambda_i$  are defined as the eigenvalues of the associated linearly symmetric transformation, and therefore, this also determines the numbers  $a_i$ . In the case of a cone (quadric of type I for  $c = 0$ ), all the numbers  $\lambda_i$  can be multiplied by a common factor that is a positive number (because of the assumptions made about  $r$ ). This means that the numbers  $a_i$  can be multiplied by an arbitrary positive common factor.

Let us note that although our line of reasoning was precisely the same as in the case of affine equivalence, the result that we obtained was different. We obtained relative to affine equivalence only a finite number of different types of inequivalent quadrics, while with respect to metric equivalence, the number is infinite: they are determined not only by a finite number of values of the index  $r$ , but also by arbitrary numbers  $a_i$  (which in the case of a cone are defined up to multiplication by a common positive factor). This fact is presented in a course in analytic geometry; for example, an ellipse with equation

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1$$



is defined by its semiaxes  $a$  and  $b$ , and if for two ellipses these are different, then the ellipses cannot be transformed into each other by a motion of the plane.

For arbitrary  $n$ , quadrics having a canonical equation (11.70) with  $k = n$  and  $c = 1$  are called *ellipsoids*. The equation of an ellipsoid can be rewritten in the form

$$\sum_{k=1}^n \left( \frac{x_i}{a_i} \right)^2 = 1, \quad (11.72)$$

from which it follows that  $|x_i/a_i| \leq 1$  and hence  $|x_i| \leq a_i$ . If the largest of these numbers  $a_1, \dots, a_n$  is denoted by  $a$ , then we obtain that  $|x_i| \leq a$ . This property is expressed by saying that the ellipsoid is a bounded set. The interested reader can easily prove that among all quadrics, only ellipsoids have this property.

If we renumber the coordinates in such a way that in the equation of the ellipsoid (11.72), the coefficients are  $a_1 \geq a_2 \geq \dots \geq a_n$ , then we obtain

$$\left( \frac{x_i}{a_1} \right)^2 \leq \left( \frac{x_i}{a_i} \right)^2 \leq \left( \frac{x_i}{a_n} \right)^2,$$

whence for every point  $\mathbf{x} = (x_1, \dots, x_n)$  lying on the ellipsoid, we have the inequality  $a_n \leq |\mathbf{x}| \leq a_1$ . This means that the distance from the center  $O$  of the ellipsoid to the point  $\mathbf{x}$  is not greater than to the point  $A = (a_1, 0, \dots, 0)$  and not less than to the point  $B = (0, \dots, 0, a_n)$ . These two points, or more precisely, the segments  $OA$  and  $OB$ , are called the *semimajor* and *semiminor* axes of the ellipsoid.

## 11.7 Quadrics in the Real Plane\*

In this section, we shall not be proving any new facts. Rather, our goal is to establish a connection between results obtained earlier with facts familiar from analytic geometry, in particular, the interpretation of quadrics in the real plane as conic sections, which was known already to the ancient Greeks.

Let us begin by considering the simplest example, in which it is possible to see the difference between the affine and projective classifications of quadrics, that is, quadrics in the real affine and real projective planes. But for this, we must first refine (or recall) the statement of the problem.

By the definition from Sect. 9.1, we may represent a projective space of arbitrary dimension  $n$  in the form  $\mathbb{P}(L)$ , where  $L$  is a vector space of dimension  $n + 1$ . An affine space of the same dimension  $n$  can be considered the affine part of  $\mathbb{P}(L)$ , determined by the condition  $\varphi \neq 0$ , where  $\varphi$  is some nonnull linear function on  $L$ . It can also be identified with the set  $W_\varphi$ , defined by the condition  $\varphi(\mathbf{x}) = 1$ . This set is an affine subspace of  $L$  (we may view  $L$  as its own space of vectors). In the sequel, we shall make use of precisely this construction of an affine space.

A quadric  $\overline{Q}$  in a projective space  $\mathbb{P}(L)$  is given by an equation  $F(\mathbf{x}) = 0$ , where  $F$  is a homogeneous second-degree polynomial. In the space  $L$ , the collection of all vectors for which  $F(\mathbf{x}) = 0$  forms a cone  $K$ . Let us recall that a *cone* is a set  $K$

such that for every vector  $\mathbf{x} \in K$ , the entire line  $\langle \mathbf{x} \rangle$  containing  $\mathbf{x}$  is also contained in  $K$ . A cone associated with a quadric is called a *quadratic cone*. From this point of view, the projective classification of quadrics coincides with the classification of quadratic cones with respect to nonsingular linear transformations.

Thus an affine quadric  $Q$  can be represented in the form  $W_\varphi \cap K$  using the previously given notation  $W_\varphi$  and  $K$ . Quadrics  $Q_1 \subset W_{\varphi_1}$  and  $Q_2 \subset W_{\varphi_2}$  are by definition affinely equivalent if there exists a nonsingular affine transformation  $W_{\varphi_1} \rightarrow W_{\varphi_2}$  mapping  $Q_1$  to  $Q_2$ . This means that we have a nonsingular linear transformation  $\mathcal{A}$  of the vector space  $L$  for which

$$\mathcal{A}(W_{\varphi_1}) = W_{\varphi_2} \quad \text{and} \quad \mathcal{A}(W_{\varphi_1} \cap K_1) = W_{\varphi_2} \cap K_2,$$

where  $K_1$  and  $K_2$  are quadratic cones associated with the quadrics  $Q_1$  and  $Q_2$ .

First of all, let us examine how the mapping  $\mathcal{A}$  acts on the set  $W_\varphi$ . To this end, let us recall that in the space  $L^*$  of linear functions on  $L$  there are defined dual transformations  $\mathcal{A}^*$  for which

$$\mathcal{A}^*(\varphi)(\mathbf{x}) = \varphi(\mathcal{A}(\mathbf{x}))$$

for all vectors  $\mathbf{x} \in L$  and  $\varphi \in L^*$ . In other words, this means that if  $\mathcal{A}^*(\varphi) = \psi$ , then the linear function  $\psi(\mathbf{x})$  is equal to  $\varphi(\mathcal{A}(\mathbf{x}))$ . Since the transformation  $\mathcal{A}$  is nonsingular, the dual transformation  $\mathcal{A}^*$  is also nonsingular, and therefore, there exists an inverse transformation  $(\mathcal{A}^*)^{-1}$ . By definition,  $(\mathcal{A}^*)^{-1}(\varphi)(\mathcal{A}(\mathbf{x})) = 1$  if  $\varphi(\mathbf{x}) = 1$ , that is,  $\mathcal{A}$  takes  $W_\varphi$  into the set  $W_{(\mathcal{A}^*)^{-1}(\varphi)}$ .

Since in previous sections, we considered only nonsingular projective quadrics, it is natural to set corresponding restrictions in the affine case as well. To this end, we shall use, as earlier, the representation of affine quadrics in the form  $Q = W_\varphi \cap K$ . A quadratic cone  $K$  determines some projection to the quadric  $\overline{Q}$ . It is easy to express this correspondence in coordinates. If we choose in  $L$  a system of coordinates  $(x_0, x_1, \dots, x_n)$ , then in  $W_{x_0}$  are defined inhomogeneous coordinates  $y_1, \dots, y_n$  by the formula  $y_i = x_i/x_0$ . If the quadric  $Q$  is given by the second-degree equation

$$f(y_1, \dots, y_n) = 0,$$

then the quadric  $\overline{Q}$  (and cone  $K$ ) is given by the equation

$$F(x_0, x_1, \dots, x_n) = 0, \quad \text{where } F = x_0^2 f\left(\frac{x_1}{x_0}, \dots, \frac{x_n}{x_0}\right).$$

Thus the projective quadric  $\overline{Q}$  is uniquely defined by the affine quadric  $Q$ .

**Definition 11.38** An affine quadric  $Q$  is said to be *nonsingular* if the associated projective quadric  $\overline{Q}$  is nonsingular.

In a space of arbitrary dimension  $n$ , all quadrics with canonical equations (11.67)–(11.69) for  $m < n$  are singular. Furthermore, a quadric of type (11.68) is

singular as well for  $m = n$ . Both these assertions can be verified directly from the definitions; we have only to designate the coordinates  $x_1, \dots, x_n$  by  $y_1, \dots, y_n$ , introduce homogeneous coordinates  $x_0 : x_1 : \dots : x_n$ , setting  $y_i = x_i/x_0$ , and multiply all the equations by  $x_0^2$ . It is very easy to write down the matrix of a quadratic form  $F(x_0, x_1, \dots, x_n)$ .

In particular, for  $n = 2$ , we obtain three equations:

$$y_1^2 + y_2^2 = 1, \quad y_1^2 - y_2^2 = 1, \quad y_1^2 + y_2 = 0. \quad (11.73)$$

From the results of Sect. 11.5, it follows that for  $n = 2$ , every nonsingular affine quadric is affinely equivalent to a quadric of one (and only one) of these three types. The corresponding quadrics are called *ellipses*, *hyperbolas*, and *parabolas*.

On the other hand, in Sect. 11.4, we saw that all nonsingular projective quadrics are projectively equivalent. This result can serve as a graphic representation of affine quadrics. As we have seen, every affine quadric can be represented in the form  $Q = W_\varphi \cap K$ , where  $K$  is some quadratic cone. It is affinely equivalent to the quadric

$$\mathcal{A}(W_\varphi \cap K) = W_{(\mathcal{A}^*)^{-1}(\varphi)} \cap \mathcal{A}(K),$$

where  $\mathcal{A}$  is an arbitrary nonsingular linear transformation of the space  $L$ .

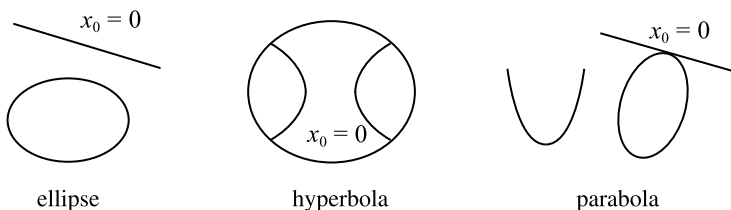
Here arises the specific nature of the case  $n = 2$  ( $\dim L = 3$ ). By what has been proved earlier, every cone  $K$  associated with a nonsingular quadric can be mapped to every other such cone by a nonsingular transformation  $\mathcal{A}$ . In particular, we may assume that  $\mathcal{A}(K) = K_0$ , where the cone  $K_0$  is given in some coordinate system  $x_0, x_1, x_2$  of the space  $L$  by the equation  $x_1^2 + x_2^2 = x_0^2$ . This cone is obtained by the rotation of one of its *generatrices*, that is, a line lying entirely on the cone (for example, the line  $x_1 = x_0, x_2 = 0$ ) about the axis  $x_0$  (that is, the line  $x_1 = x_2 = 0$ ). In the cone  $K_0$  that we have chosen, the angle between the generatrix and the axis  $x_0$  is equal to  $\pi/4$ . In other words, this means that each pole of the cone  $K_0$  is obtained by a rotation of the sides of an isosceles right triangle around its bisector.

Setting  $(\mathcal{A}^*)^{-1}(\varphi) = \psi$ , we obtain that an arbitrary nonsingular affine quadric is affinely equivalent to the quadric  $W_\psi \cap K_0$ . Here  $W_\psi$  is an arbitrary plane in the space  $L$  not passing through the vertex of the cone  $K_0$ , that is, through the point  $O = (0, 0, 0)$ . Thus every nonsingular affine quadric is affinely equivalent to a planar section of a right circular cone. This explains the terminology *conic* used for quadrics in the plane.

It is well known from analytic geometry how the three conics that we have found (ellipses, hyperbolas, and parabolas) are obtained from a single (from the point of view of projective classification) curve. If we begin with equations (11.73), then the difference in the three types is revealed by writing these equations in homogeneous coordinates. Setting  $y_1 = x_1/x_0$  and  $y_2 = x_2/x_0$ , we obtain the equations

$$x_1^2 + x_2^2 = x_0^2, \quad x_1^2 - x_2^2 = x_0^2, \quad x_1^2 - x_0x_2 = 0. \quad (11.74)$$

The differences among these equations can be found in the different natures of the sets of intersection with the infinite line  $l_\infty$  given by the equation  $x_0 = 0$ . For an



**Fig. 11.7** Intersection of a conic with an infinite line

ellipse, this set is empty; for a hyperbola, it consists of two points,  $(0 : 1 : 1)$  and  $(0 : 1 : -1)$ , and for a parabola, it consists of the single point  $(0 : 0 : 1)$  (substitution into equation (11.73) shows that the line  $l_\infty$  is tangent to the parabola at the point of intersection); see Fig. 11.7.

We saw in Sect. 9.2 that an affine transformation coincides with a projective transformation that preserves the line  $l_\infty$ . Therefore, the type of set  $\overline{Q} \cap l_\infty$  (empty set, two points, one point) should be the same for affinely equivalent quadrics  $Q$ . In our case, the actual content of what we proved in Sect. 11.4 is that the type of set  $\overline{Q} \cap l_\infty$  determines the quadric  $Q$  up to affine equivalence.

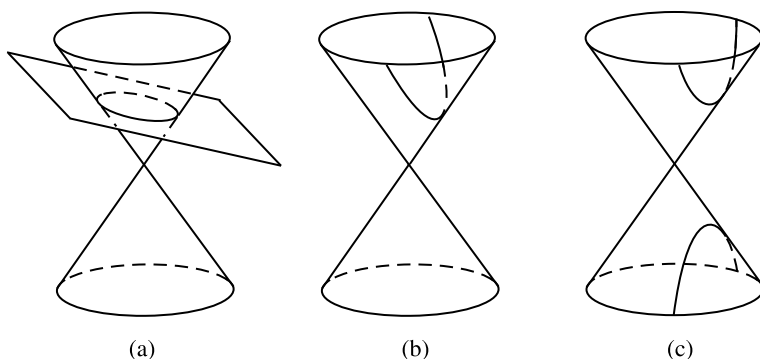
But if we begin with the representation of a conic as the intersection of the cone  $K_0$  with the plane  $W_\psi$ , then different types appear due to a different disposition of the plane  $W_\psi$  with respect to the cone  $K_0$ . Let us recall that the vertex  $O$  of the cone  $K_0$  partitions it into two poles. If the equation of the cone has the form  $x_1^2 + x_2^2 = x_0^2$ , then each pole is determined by the sign of  $x_0$ .

Let us denote by  $L_\psi$  the plane parallel to  $W_\psi$  and passing through the point  $O$ . This plane is given by the equation  $\psi = 0$ . If  $L_\psi$  has no points of intersection with the cone  $K_0$  other than  $O$ , then  $W_\psi$  intersects one of its poles (for example, the one within which lie the point of intersection  $W_\psi$  and the axis  $x_0$ ). In this case, the conic  $W_\psi \cap K_0$  lies within one pole and is an ellipse.

For example, in the special case in which the plane  $W_\psi$  is orthogonal to the axis  $x_0$ , we obtain a circle. If we move the plane  $W_\psi$  (for example, decrease its angle with the axis  $x_0$ ), then in its intersection with the cone  $K_0$ , an ellipse is obtained whose eccentricity increases as the angle is decreased; see Fig. 11.8(a). The limiting position is reached when the plane  $L_\psi$  is tangent to the cone  $K_0$  on a generatrix. Then  $W_\psi$  again intersects in one pole (the one that contains the intersection with the axis  $x_0$ ). This intersection is a parabola; see Fig. 11.8(b). And if the plane  $L_\psi$  intersects  $K_0$  in two different generatrices, then  $W_\psi$  intersects both of its poles (on the side of the plane  $L_\psi$  on which is located the plane  $W_\psi$  parallel to it). This intersection is a hyperbola; see Fig. 11.8(c).

The connection between planar quadrics and conic sections is revealed particularly clearly by the metric classification of such quadrics, which forms part of any sufficiently rigorous course in analytic geometry. Let us recall only the main results.

As was done in Sect. 11.5, we must exclude from consideration those conics that are cylinders and those that are unions of vector subspaces (that is, in our case, lines or points). Then the results obtained in Sect. 11.5 give us (in coordinates  $x, y$ ) the



**Fig. 11.8** Conic sections

following three types of conic:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \quad x^2 + a^2y = 0, \quad (11.75)$$

where  $a > 0$  and  $b > 0$ . From the point of view of affine classification presented above, curves of the first type are ellipses, those of the second type are hyperbolas, and those of the third type are parabolas.

Let us recall that in a course in analytic geometry, these curves are defined as *geometric loci* of points of the plane satisfying certain conditions. Namely, an ellipse is the geometric locus of points the sum of whose distances from two given points in the plane is constant. A hyperbola is defined analogously with sum replaced by difference. A parabola is the geometric locus of points equidistant from a given point and a given line that does not pass through the given point.

There is an elegant and elementary proof of the fact that all ellipses, hyperbolas, and parabolas are not only affinely, but also *metrically*, that is, as geometric loci of points, equivalent to planar sections of a right circular cone. Let us recall that by *right circular cone* we mean a cone  $K$  in three-dimensional space obtained as the result of a rotation of a line about some other line, called the *axis* of the cone. The lines forming the cone are called its *generatrices*; they intersect the axis of the cone in one common point, called its *vertex*.

In other words, this result means that the section of a right circular cone with a plane not passing through the vertex of the cone is either an ellipse, a hyperbola, or a parabola, and every ellipse, hyperbola, and parabola coincides with the intersection of a right circular cone with a suitable plane.<sup>5</sup>

<sup>5</sup>The proof of this fact is due to the Franco-Belgian mathematician Germinal Pierre Dandelin. It can be found, for example, in A.P. Veselov and E.V. Troitsky, *Lectures in Analytic Geometry* (in Russian); B.N. Delone and D.A. Raikov, *Analytic Geometry* (in Russian); P. Dandelin, *Mémoire sur l'hyperboloïde de révolution, et sur les hexagones de Pascal et de M. Brianchon*; D. Hilbert and S. Cohn-Vossen, *Geometry and the Imagination*.

## Chapter 12

# Hyperbolic Geometry

The discovery of hyperbolic (or Lobachevskian) geometry had an enormous impact on the development of mathematics and on how the relationship between mathematics and the real world was understood. The discussions that swirled around the new geometry also seem to have influenced the views of many in the humanities, who, in this regard, unfortunately were too much taken by a literary image: the contrast between “down-to-earth” Euclidean geometry and the “otherworldly” non-Euclidean geometry invented by learned mathematicians. It seemed that the difference between the two geometries was that in the first geometry, as was clear to everyone, parallel lines did not intersect, while in the second, what to normal intelligence was difficult of comprehension, they do intersect. However, of course, this is exactly the opposite of the truth: in the non-Euclidean geometry of Lobachevsky, given a point external to a given line, it is possible for *infinitely many* lines to pass through the point without intersecting the line. It is this that distinguishes Lobachevsky’s geometry from that of Euclid.

Ivan Karamazov, in Dostoevsky’s novel *The Brothers Karamazov*, likely sowed confusion among those in the humanities with the following literary image:

At the same time there were and are even now geometers and philosophers, even some of the most outstanding among them, who doubt that the whole universe, or, even more broadly, the whole of being, was created purely in accordance with Euclidean geometry; they even dare to dream that two parallel lines, which according to Euclid cannot possibly meet on earth, may perhaps meet somewhere in infinity.

Around the time this novel was being written, Friedrich Engels wrote *Anti-Dühring*, where an even more vivid image is used:

But in higher mathematics, another contradiction is achieved, that lines that intersect before our eyes, nevertheless a mere five or six centimeters from their point of intersection are to be considered parallel, that is, lines that cannot intersect even when extended to infinity.

In this, the author sees the manifestation of some sort of “dialectic.”

And even up to the present, it is possible to encounter, in print, such literary images that oppose Euclidean and non-Euclidean geometries by saying that in the former, parallel lines do not intersect, while in the latter, they “intersect somewhere

or other.” Usually, by non-Euclidean geometry is meant the hyperbolic geometry of Lobachevsky, which is quite understandable by anyone who has passed a college course in some technical subject, and there are many such people today. To be sure, nowadays, this is presented in mathematics departments in more advanced courses in differential geometry. But hyperbolic geometry is so tightly linked to a first course in linear algebra, that it would be a pity not to say something about it here.

## 12.1 Hyperbolic Space\*

In this chapter we shall be dealing exclusively with *real* vector spaces.

We shall define *hyperbolic space* of dimension  $n$ , which we shall hereinafter denote by  $\mathbb{L}_n$  or simply  $\mathbb{L}$  if we do not need to indicate the dimension, as a part of  $n$ -dimensional projective space  $\mathbb{P}(\mathbb{L})$ , where  $\mathbb{L}$  is a real vector space of dimension  $n + 1$ . We shall denote the dimension of the space  $\mathbb{L}$  by  $\dim \mathbb{L}$ .

Let us equip  $\mathbb{L}$  with a pseudo-Euclidean product  $(\mathbf{x}, \mathbf{y})$ ; see Sect. 7.7. Let us recall that there, the quadratic form  $(\mathbf{x}^2)$  has index of inertia  $n$ , and in some basis  $\mathbf{e}_1, \dots, \mathbf{e}_{n+1}$  (called orthonormal) for the vector

$$\mathbf{x} = \alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n + \alpha_{n+1} \mathbf{e}_{n+1}, \quad (12.1)$$

it takes the form

$$(\mathbf{x}^2) = \alpha_1^2 + \dots + \alpha_n^2 - \alpha_{n+1}^2. \quad (12.2)$$

In the pseudo-Euclidean space  $\mathbb{L}$ , let us consider the light cone  $V$  defined by the condition  $(\mathbf{x}^2) = 0$ . We say that a vector  $\mathbf{a}$  lies *inside* the cone  $V$  if  $(\mathbf{a}^2) < 0$  (recall that in Chap. 7, we called such vectors timelike). It is obvious that the same then holds as well for all vectors on the line  $\langle \mathbf{a} \rangle$ , since  $((\mathbf{a}\mathbf{a})^2) = \alpha^2 (\mathbf{a}^2) < 0$ , and we shall consider this space over the field of real numbers. Such lines are also said to lie *inside* the light cone  $V$ .

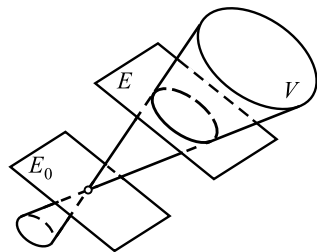
Points of the projective space  $\mathbb{P}(\mathbb{L})$  corresponding to lines of the space  $\mathbb{L}$  lying inside the light cone  $V$  are called *points* of the space  $\mathbb{L}$ . Consequently, they correspond to those lines  $\langle \mathbf{x} \rangle$  of the space  $\mathbb{L}$  that in the form (12.1) satisfy the inequality

$$\alpha_1^2 + \dots + \alpha_n^2 < \alpha_{n+1}^2. \quad (12.3)$$

In view of condition (12.3), the set  $\mathbb{L} \subset \mathbb{P}(\mathbb{L})$  is contained in *one* affine subset  $\alpha_{n+1} \neq 0$  (see Sect. 9.1). Indeed, in the case  $\alpha_{n+1} = 0$ , we would obtain in (12.3) the inequality  $\alpha_1^2 + \dots + \alpha_n^2 < 0$ , which is impossible in view of the fact that  $\alpha_1, \dots, \alpha_n$  are real. As we did previously in Sect. 9.1, we can identify the affine subset  $\alpha_{n+1} \neq 0$  with the affine subspace  $E : \alpha_{n+1} = 1$  and hence view  $\mathbb{L}$  as a part of  $E$ ; see Fig. 12.1.

The space of vectors of the affine space  $E$  is the vector subspace  $E_0 \subset \mathbb{L}$  defined by the condition  $\alpha_{n+1} = 0$ . In other words,  $E_0 = \langle \mathbf{e}_1, \dots, \mathbf{e}_n \rangle$ . Let us note that the space of vectors  $E_0$  is not simply a vector space. As a subspace of the pseudo-Euclidean space  $\mathbb{L}$ , it would seem that it should also be a pseudo-Euclidean space.

**Fig. 12.1** Model of hyperbolic space



But in fact, as can be seen from formula (12.2), the inner product  $(\mathbf{x}, \mathbf{y})$  makes it a *Euclidean* space, in which the vectors  $\mathbf{e}_1, \dots, \mathbf{e}_n$  form an orthonormal basis. This means that  $E$  is an affine Euclidean space, and the basis  $\mathbf{e}_1, \dots, \mathbf{e}_{n+1}$  of the space  $L$  forms within it a frame of reference with respect to which a point of the hyperbolic space  $\mathbb{L} \subset E$  with coordinates  $(y_1, \dots, y_n)$  is characterized by the relationship

$$y_1^2 + \dots + y_n^2 < 1, \quad y_i = \frac{\alpha_i}{\alpha_{n+1}}, i = 1, \dots, n. \quad (12.4)$$

This set is called the *interior* of the unit sphere in  $E$  and will be denoted by  $U$ .

Let us now turn our attention to identifying the subspaces of a hyperbolic space. They correspond to those vector spaces  $L' \subset L$  that have a common point with the interior of the light cone  $V$ , that is, they contain a timelike vector  $\mathbf{a} \in L'$ . The inner product  $(\mathbf{x}, \mathbf{y})$  defined in  $L$  is clearly also defined for all vectors in the subspace  $L' \subset L$ . The space  $L'$  contains the timelike vector  $\mathbf{a}$ , and therefore, by Lemma 7.53, it is a pseudo-Euclidean space, and therefore, the associated hyperbolic space  $\mathbb{L}' \subset \mathbb{P}(L')$  is defined. Since  $\mathbb{P}(L') \subset \mathbb{P}(L)$  is a projective subspace, it follows that  $\mathbb{L}' \subset \mathbb{P}(L)$ . But hyperbolic space  $\mathbb{L}'$  is defined by the condition  $(\mathbf{x}^2) < 0$  both in  $\mathbb{P}(L)$  and in  $\mathbb{P}(L')$ , and therefore,  $\mathbb{L}' \subset \mathbb{L}$ . Here by definition,  $\dim \mathbb{L}' = \dim \mathbb{P}(L') = \dim L' - 1$ . The hyperbolic space  $\mathbb{L}'$  thus constructed is called a *subspace* in  $\mathbb{L}$ .

In particular, if  $L'$  is a hyperplane in  $L$ , then  $\dim \mathbb{L}' = \dim \mathbb{L} - 1$ , and then the subspace  $\mathbb{L}' \subset \mathbb{L}$  is called a *hyperplane* in  $\mathbb{L}$ .

In the sequel we shall require the partition of  $\mathbb{L}$  into two parts by the hyperplane  $\mathbb{L}' \subset \mathbb{L}$ :

$$\mathbb{L} \setminus \mathbb{L}' = \mathbb{L}^+ \cup \mathbb{L}^-, \quad \mathbb{L}^+ \cap \mathbb{L}^- = \emptyset, \quad (12.5)$$

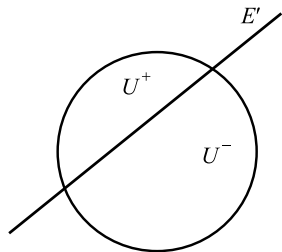
similar to how in Sect. 3.2, the partition of the vector space  $L$  into two half-spaces was accomplished with the help of the hyperplane  $L' \subset L$ .

The partition (12.5) of the space  $\mathbb{L}$  cannot be accomplished by an analogous partition of the projective space  $\mathbb{P}(L)$ . Indeed, if we use the definition of the subsets  $L^+$  and  $L^-$  from Sect. 3.2, then we see that for a vector  $\mathbf{x} \in L^+$ , the vector  $\alpha \mathbf{x}$  is in  $L^-$  if  $\alpha < 0$ , so that the condition  $\mathbf{x} \in L^+$  does not hold for the line  $\langle \mathbf{x} \rangle$ . But such a partition is possible for the affine Euclidean space  $E$ ; it was constructed in Sect. 8.2 (see p. 299).

Let us recall that the partition of the affine space  $E$  by the hyperplane  $E' \subset E$  was defined via the partition of the space of vectors  $E_0$  of the affine space  $E$  with



**Fig. 12.2** Hyperbolic half-spaces



the aid of the hyperplane  $E'_0 \subset E_0$  corresponding to the affine hyperplane  $E'$ , that is, consisting of vectors  $\overrightarrow{AB}$ , where  $A$  and  $B$  are all possible points of  $E'$ . If we are given a partition  $E_0 \setminus E'_0 = E_0^+ \cup E_0^-$ , then we must choose an arbitrary point  $O \in E'$  and define  $E^+$  as the collection of all points  $A \in E$  such that  $\overrightarrow{OA} \in E_0^+$  ( $E^-$  is defined analogously). The sets  $E^+$  and  $E^-$  thus obtained are called *half-spaces*, and they do not depend on the choice of point  $O \in E'$ . Thus we have partitioned the set  $E \setminus E'$  into two half-spaces:  $E \setminus E' = E^+ \cup E^-$ .

Let  $L'$  be a hyperplane in the pseudo-Euclidean space  $L$  having nonempty intersection with the interior of the light cone  $V$ , and let  $E'$  be the associated hyperplane in the affine space  $E$ , that is,  $E' = E \cap \mathbb{P}(L')$ . Then  $E'$  has nonempty intersection with the interior of the unit sphere  $U$ , given by relationship (12.4), and for the set  $L \subset E$ , we obtain the partition (12.5), where

$$L' = L \cap E', \quad L^+ = E^+ \cap L, \quad L^- = E^- \cap L. \quad (12.6)$$

The sets  $L^+$  and  $L^-$  defined by relationships (12.6) are called *half-spaces* of the space  $L$ .

To put it more simply, the hyperplane  $E'$  divides the interior of the sphere  $U \subset E$  identified with the space  $L$  into two parts,  $U^+$  and  $U^-$  (see Fig. 12.2), which correspond to the half-spaces  $L^+$  and  $L^-$ .

Let us show that both half-spaces  $L^+$  and  $L^-$  are nonempty, although Fig. 12.2 is sufficiently convincing by itself. We give the proof for  $L^+$  (for  $L^-$ , the proof is similar).

Let us consider an arbitrary point  $O \in E' \cap L$ . It corresponds to the vector  $\mathbf{a} = \alpha_1 \mathbf{e}_1 + \cdots + \alpha_n \mathbf{e}_n + \mathbf{e}_{n+1}$  with  $(\mathbf{a}^2) < 0$  (see the definition of the affine space  $E$  on p. 434). Let  $\mathbf{c} \in E_0^+$  and  $B \in E^+$  be points such that  $\overrightarrow{OB} = \mathbf{c}$ . Let us consider vectors  $\mathbf{b}_t = \mathbf{a} + t\mathbf{c} \in L$  and points  $B_t \in E$  for which  $\overrightarrow{OB_t} = \mathbf{b}_t$  for varying values of  $t \in \mathbb{R}$ . Let us note that if  $t > 0$ , then  $B_t \in E^+$ , and if here  $(\mathbf{b}_t^2) < 0$ , then  $B_t \in E^+ \cap L = L^+$ . As can be seen without difficulty, the scalar square  $(\mathbf{b}_t^2)$  is a quadratic trinomial in  $t$ :

$$(\mathbf{b}_t^2) = ((\mathbf{a} + t\mathbf{c})^2) = (\mathbf{a}^2) + 2t(\mathbf{a}, \mathbf{c}) + t^2(\mathbf{c}^2) = P(t). \quad (12.7)$$

By our selection, the vector  $\mathbf{c} \neq 0$  belongs to the Euclidean space  $E_0$ , and therefore,  $(\mathbf{c}^2) > 0$ . On the other hand, by assumption, we have  $(\mathbf{a}^2) < 0$ . This yields that the discriminant of the quadratic trinomial  $P(t)$  on the right-hand side of relationship (12.7) is positive, and therefore,  $P(t)$  has two real roots,  $t_1$  and  $t_2$ , and from the

condition  $(a^2) < 0$  it follows that they have different signs, that is,  $t_1 t_2 < 0$ . Then, as is easy to see,  $P(t) < 0$  for every  $t$  between the roots  $t_1$  and  $t_2$ . We will choose a positive such number  $t$ .

Since the hyperbolic space  $\mathbb{L}$  can be viewed as a part of the affine space  $E$ , then from  $E$  we can transfer onto  $\mathbb{L}$  the notion of line segment, the notion of *lying between* for three points on a line segment, and the notion of convexity. An easy verification (analogous to what we did at the end of Sect. 8.2) shows that the subsets  $\mathbb{L}^+$  and  $\mathbb{L}^-$  introduced earlier of the set  $\mathbb{L} \setminus \mathbb{L}'$  are characterized by the property of convexity: if two points  $A, B$  are in  $\mathbb{L}^+$ , then all points lying on the segment  $[A, B]$  are also in  $\mathbb{L}^+$  (the same clearly holds for the subset  $\mathbb{L}^-$ ).

Let us consider linear transformations  $\mathcal{A}$  of a vector space  $L$  that are Lorentz transformations with respect to a symmetric bilinear form  $\varphi(x, y)$  corresponding to the quadratic form  $(x^2)$  and the associated projective transformations  $\mathbb{P}(\mathcal{A})$ . The latter transformations obviously take the set  $\mathbb{L}$  to itself: given that a transformation  $\mathcal{A}$  is a Lorentz transformation and from the condition  $(x^2) < 0$ , it follows that  $(\mathcal{A}(x)^2) = (x^2) < 0$ . The transformations of the set  $\mathbb{L}$  that arise in this way are called *motions* of the hyperbolic space  $\mathbb{L}$ .

Thus motions of the space  $\mathbb{L}$  are projective transformations of the projective space  $\mathbb{P}(L)$  containing  $\mathbb{L}$  and taking the quadratic form  $(x^2)$  into itself. By what we have said thus far, the definition of the interior of the light cone  $V$  can be written in homogeneous coordinates in the form

$$x_1^2 + \cdots + x_n^2 - x_{n+1}^2 < 0, \quad (12.8)$$

and in inhomogeneous coordinates  $y_i = x_i/x_{n+1}$  in the form

$$y_1^2 + \cdots + y_n^2 < 1. \quad (12.9)$$

We consider motions of a hyperbolic space as transformations of the set  $\mathbb{L}$ , that is, as transformations taking the interior of the unit sphere given by condition (12.9) into itself.

Let us write down some simple properties of motions:

*Property 12.1* The sequential application (composition) of two motions  $f_1$  and  $f_2$  (as transformations of the set  $\mathbb{L}$ ) is again a motion.

This follows at once from the fact that the composition of nonsingular transformations  $\mathcal{A}_1$  and  $\mathcal{A}_2$  is a nonsingular transformation, and this holds as well for the corresponding projective transformations  $\mathbb{P}(\mathcal{A}_1)$  and  $\mathbb{P}(\mathcal{A}_2)$ . Moreover, if  $\mathcal{A}_1$  and  $\mathcal{A}_2$  are Lorentz transformations with respect to the bilinear form  $\varphi(x, y)$ , then the result of their composition has the same property.

*Property 12.2* A motion is a bijection of  $\mathbb{L}$  to itself.

This assertion follows from the fact that the corresponding transformations  $\mathcal{A} : L \rightarrow L$  and  $\mathbb{P}(\mathcal{A}) : \mathbb{P}(L) \rightarrow \mathbb{P}(L)$  are bijections. But by the definition of a hyperbolic

space, it is also necessary to verify that every line contained in the interior of the light cone  $V$  is the image of a similar such line. If we have the line  $\langle \mathbf{a} \rangle$  with a timelike vector  $\mathbf{a}$ , then we know already that there exists a vector  $\mathbf{b}$  such that  $\mathcal{A}(\mathbf{b}) = \mathbf{a}$ . Since  $\mathcal{A}$  is a Lorentz transformation of a pseudo-Euclidean space  $L$ , we have the relationship  $(\mathbf{b}^2) = (\mathcal{A}(\mathbf{b})^2) = (\mathbf{a}^2) < 0$ , from which it follows that the vector  $\mathbf{b}$  is also timelike. Thus the transformation  $\mathcal{A}$  takes the line  $\langle \mathbf{b} \rangle$  lying inside  $V$  into the line  $\langle \mathbf{a} \rangle$ , also inside  $V$ .

*Property 12.3* Like every bijection, a motion  $f$  has an inverse transformation  $f^{-1}$ . It is also a motion.

The verification of this property is trivial.

At first glance, it is not obvious that there are “sufficiently many” motions of a hyperbolic space. We shall establish this a bit later, but for now, we shall point out some important types of motions.

A transformation  $g$  is of type (a) if  $g = \mathbb{P}(\mathcal{A})$ , where  $\mathcal{A}$  is a Lorentz transformation of the space  $L$  such that  $\mathcal{A}(\mathbf{e}_{n+1}) = \mathbf{e}_{n+1}$ .

Since the basis  $\mathbf{e}_1, \dots, \mathbf{e}_{n+1}$  of the pseudo-Euclidean space  $L$  is orthonormal, we have the decomposition

$$L = \langle \mathbf{e}_{n+1} \rangle \oplus \langle \mathbf{e}_{n+1} \rangle^\perp, \quad \langle \mathbf{e}_{n+1} \rangle^\perp = \langle \mathbf{e}_1, \dots, \mathbf{e}_n \rangle, \quad (12.10)$$

and all transformations  $\mathcal{A} : L \rightarrow L$  with the indicated property take the subspace  $E_0 = \langle \mathbf{e}_1, \dots, \mathbf{e}_n \rangle$  into itself.

Conversely, if we define  $\mathcal{A} : L \rightarrow L$  as an orthogonal transformation of the Euclidean subspace  $E_0$  and set  $\mathcal{A}(\mathbf{e}_{n+1}) = \mathbf{e}_{n+1}$ , then  $\mathbb{P}(\mathcal{A})$  will of course be a motion of the hyperbolic space. In other words, these transformations can be described as orthogonal transformations of inhomogeneous coordinates. All thus constructed motions of the space  $\mathbb{L}$  have the fixed point  $O$  corresponding to the line  $\langle \mathbf{e}_{n+1} \rangle$  in  $L$ , or in other words, the point  $O = (0, \dots, 0)$  in the inhomogeneous system of coordinates  $(y_1, \dots, y_n)$ .

From the point of view of hyperbolic space, the constructed motions precisely *coincide* with those motions that leave the point  $O \in \mathbb{L}$  fixed. Indeed, as we have seen, the point  $O$  corresponds to the line  $\langle \mathbf{e}_{n+1} \rangle$ , and the motion  $g$  is equal to  $\mathbb{P}(\mathcal{A})$ , where  $\mathcal{A}$  is a Lorentz transformation of the space  $L$ . The condition  $g(O) = O$  means that  $\mathcal{A}(\langle \mathbf{e}_{n+1} \rangle) = \langle \mathbf{e}_{n+1} \rangle$ , that is,  $\mathcal{A}(\mathbf{e}_{n+1}) = \lambda \mathbf{e}_{n+1}$ . From the fact that  $\mathcal{A}$  is a Lorentz transformation, it follows that  $\lambda = \pm 1$ . By multiplying  $\mathcal{A}$  by  $\pm 1$ , which obviously does not change the transformation  $g = \mathbb{P}(\mathcal{A})$ , we can obtain that the conditions  $\mathcal{A}(\mathbf{e}_{n+1}) = \mathbf{e}_{n+1}$  are satisfied, whence by definition, it follows that  $g$  is a transformation of type (a).

Type (b) is connected with a certain line  $\mathbb{L}_1 \subset \mathbb{L}$  of a hyperbolic space. By definition, the line  $\mathbb{L}_1$  is determined by the plane  $L' \subset L$ ,  $\dim L' = 2$ . Since by assumption, the plane  $L'$  must contain at least one timelike vector  $\mathbf{x}$ , it follows by Lemma 7.53 (p. 271) that it is a pseudo-Euclidean space. From formula (6.28) and Theorem 6.17

(law of inertia), it follows that all such spaces of a given dimension are isomorphic. Therefore, we can choose a basis in  $L'$  with any convenient Gram matrix, provided only that it defines a pseudo-Euclidean plane. We have seen (in Example 7.49, p. 269) that it is convenient to choose as such a basis the lightlike vectors  $f_1, f_2$ , for which

$$(f_1^2) = (f_2^2) = 0, \quad (f_1, f_2) = \frac{1}{2},$$

and this means that for every vector  $x = x f_1 + y f_2$ , its scalar square ( $x^2$ ) is equal to  $xy$ . In Example 7.61 (p. 277), we found explicit formulas for the Lorentz transformations of a pseudo-Euclidean plane in such a basis:

$$\mathcal{U}(f_1) = \alpha f_1, \quad \mathcal{U}(f_2) = \alpha^{-1} f_2 \quad (12.11)$$

or

$$\mathcal{U}(f_1) = \alpha f_2, \quad \mathcal{U}(f_2) = \alpha^{-1} f_1, \quad (12.12)$$

where  $\alpha$  is an arbitrary nonzero number. In the sequel we shall need only transformations given by formula (12.11).

Since  $L'$  is a nondegenerate space, it follows that by Theorem 6.9, we have the decomposition  $L = L' \oplus (L')^\perp$ . Let us now define a linear transformation  $\mathcal{A}$  of the space  $L$  by the condition

$$\mathcal{A}(x + y) = \mathcal{U}(x) + y, \quad \text{where } x \in L', y \in (L')^\perp, \quad (12.13)$$

where  $\mathcal{U}$  is one of the Lorentz transformations of the pseudo-Euclidean plane  $L'$  defined by formulas (12.11) and (12.12). It is clear that then  $\mathcal{A}$  is a Lorentz transformation of the space  $L$ .

A motion of type (b) of the space  $L$  is a transformation  $\mathbb{P}(\mathcal{A})$  obtained in the case that in formula (12.13), we take as  $\mathcal{U}$  the transformation given by relationships (12.11). All motions thus constructed have a *fixed line*  $L_1$  corresponding to the plane  $L'$ .

It is quite obvious that motions of types (a) and (b) do not exhaust all motions of the hyperbolic plane, even if in the definition of motions of type (b), as  $\mathcal{U}$  in formula (12.13) we were to use transformations  $\mathcal{U}$  given not only by relationships (12.11), but also by (12.12). For example, they certainly do not include motions associated with Lorentz transformations that have a three-dimensional cyclic subspace (see Corollary 7.66 and Example 7.67). However, for our further purposes, it will suffice to use only motions of these two types.

*Example 12.4* In the sequel we are going to require explicit formulas for transformations of type (b) in the case of the hyperbolic plane (that is, for  $n = 2$ ). In this case,  $L$  is a three-dimensional pseudo-Euclidean space, and in the orthonormal basis  $e_1, e_2, e_3$ , such that

$$(e_1^2) = 1, \quad (e_2^2) = 1, \quad (e_3^2) = -1,$$

the scalar square of the vector  $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3$  is equal to  $(\mathbf{x}^2) = x_1^2 + x_2^2 - x_3^2$ . The points of the hyperbolic plane  $\mathbb{L}$  are contained in the affine plane  $x_3 = 1$ , have inhomogeneous coordinates  $x = x_1/x_3$  and  $y = x_2/x_3$ , and satisfy the relationship  $x^2 + y^2 < 1$ .

For writing the transformation  $\mathcal{A}$ , let us consider the pseudo-Euclidean plane  $\mathbf{L}' = \langle \mathbf{e}_1, \mathbf{e}_3 \rangle$  and let us choose in it a basis consisting of lightlike vectors  $\mathbf{f}_1, \mathbf{f}_2$  associated with vectors  $\mathbf{e}_1, \mathbf{e}_3$  by the relationships

$$\mathbf{f}_1 = \frac{\mathbf{e}_1 + \mathbf{e}_3}{2}, \quad \mathbf{f}_2 = \frac{\mathbf{e}_1 - \mathbf{e}_3}{2}, \quad (12.14)$$

from which we also obtain the inverse formulas  $\mathbf{e}_1 = \mathbf{f}_1 + \mathbf{f}_2$  and  $\mathbf{e}_3 = \mathbf{f}_1 - \mathbf{f}_2$ .

Let us note that the orthogonal complement  $(\mathbf{L}')^\perp$  equals  $\langle \mathbf{e}_2 \rangle$ , and by Theorem 6.9, we have the decomposition  $\mathbf{L} = \mathbf{L}' \oplus \langle \mathbf{e}_2 \rangle$ . Then in accord with formula (12.13), for the vector  $\mathbf{z} = \mathbf{x} + \mathbf{y}$ , where  $\mathbf{x} \in \mathbf{L}'$  and  $\mathbf{y} \in \langle \mathbf{e}_2 \rangle$ , we obtain the value  $\mathcal{A}(\mathbf{z}) = \mathcal{U}(\mathbf{x}) + \mathbf{y}$ , where  $\mathcal{U} : \mathbf{L}' \rightarrow \mathbf{L}'$  is the Lorentz transformation defined in the basis  $\mathbf{f}_1, \mathbf{f}_2$  by formula (12.11). From this, taking into account expression (12.14), we obtain

$$\mathcal{U}(\mathbf{e}_1) = \frac{\alpha + \alpha^{-1}}{2}\mathbf{e}_1 + \frac{\alpha - \alpha^{-1}}{2}\mathbf{e}_3, \quad \mathcal{U}(\mathbf{e}_3) = \frac{\alpha - \alpha^{-1}}{2}\mathbf{e}_1 + \frac{\alpha + \alpha^{-1}}{2}\mathbf{e}_3.$$

Let us set

$$a = \frac{\alpha + \alpha^{-1}}{2}, \quad b = \frac{\alpha - \alpha^{-1}}{2}. \quad (12.15)$$

Then  $a + b = \alpha$  and  $a^2 - b^2 = 1$ . It is obvious that any numbers  $a$  and  $b$  satisfying these relationships can be defined in terms of the number  $\alpha = a + b$  by formulas (12.15). Therefore, we obtain the linear transformation  $\mathcal{A} : \mathbf{L} \rightarrow \mathbf{L}$ , for which

$$\mathcal{A}(\mathbf{e}_1) = a\mathbf{e}_1 + b\mathbf{e}_3, \quad \mathcal{A}(\mathbf{e}_2) = \mathbf{e}_2, \quad \mathcal{A}(\mathbf{e}_3) = b\mathbf{e}_1 + a\mathbf{e}_3.$$

It is easy to see that for such a transformation, the vector  $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3$  is carried to the vector

$$\mathcal{A}(\mathbf{x}) = (ax_1 + bx_3)\mathbf{e}_1 + x_2\mathbf{e}_2 + (bx_1 + ax_3)\mathbf{e}_3.$$

In inhomogeneous coordinates,  $x = x_1/x_3$  and  $y = x_2/x_3$ . This means that a point with coordinates  $(x, y)$  is carried to the point with coordinates  $(x', y')$ , where

$$x' = \frac{ax + b}{bx + a}, \quad y' = \frac{y}{bx + a}, \quad a^2 - b^2 = 1. \quad (12.16)$$

This particular type of motion yields, however, an important general property:

**Theorem 12.5** *For every pair of points of a hyperbolic space there exists a motion taking one point into the other.*

*Proof* Let the first point correspond to the line  $\langle \mathbf{a} \rangle$ , and the second to the line  $\langle \mathbf{b} \rangle$ , where  $\mathbf{a}, \mathbf{b} \in \mathbb{L}$ . If the vectors  $\mathbf{a}$  and  $\mathbf{b}$  are proportional, that is,  $\langle \mathbf{a} \rangle = \langle \mathbf{b} \rangle$ , then our requirements will be satisfied by the identity transformation of the space  $\mathbb{L}$  (which can be obtained in the form  $\mathbb{P}(\mathcal{E})$ , where  $\mathcal{E}$  is the identity transformation of the space  $\mathbb{L}$ ).

But if  $\langle \mathbf{a} \rangle \neq \langle \mathbf{b} \rangle$ , that is,  $\dim \langle \mathbf{a}, \mathbf{b} \rangle = 2$ , then let us set  $\mathbb{L}' = \langle \mathbf{a}, \mathbf{b} \rangle$ . Let us consider the Lorentz transformation  $\mathcal{U} : \mathbb{L}' \rightarrow \mathbb{L}'$  of type (b) given by formula (12.11), the corresponding Lorentz transformation  $\mathcal{A} : \mathbb{L} \rightarrow \mathbb{L}$  defined by formula (12.13), and the projective transformation  $\mathbb{P}(\mathcal{A}) : \mathbb{P}(\mathbb{L}) \rightarrow \mathbb{P}(\mathbb{L})$ .

Let us show that the constructed projective transformation  $\mathbb{P}(\mathcal{A})$  takes a point corresponding to the line  $\langle \mathbf{a} \rangle$  to a point corresponding to the line  $\langle \mathbf{b} \rangle$ , that is, the linear transformation  $\mathcal{A} : \mathbb{L} \rightarrow \mathbb{L}$  takes the line  $\langle \mathbf{a} \rangle$  to the line  $\langle \mathbf{b} \rangle$ . Since vectors  $\mathbf{a}$  and  $\mathbf{b}$  are contained in the plane  $\mathbb{L}'$ , then by definition, it suffices for us to prove that for an appropriate choice of number  $\alpha$ , the transformation  $\mathcal{U} : \mathbb{L}' \rightarrow \mathbb{L}'$  given by formula (12.11) takes the line  $\langle \mathbf{a} \rangle$  to the line  $\langle \mathbf{b} \rangle$ .

This is easily verified by a simple calculation using the basis  $\mathbf{f}_1, \mathbf{f}_2$ , given by formula (12.14), in the pseudo-Euclidean plane  $\mathbb{L}'$ . Let us consider the timelike vectors  $\mathbf{a} = a_1 \mathbf{f}_1 + a_2 \mathbf{f}_2$  and  $\mathbf{b} = b_1 \mathbf{f}_1 + b_2 \mathbf{f}_2$ . Since in the chosen basis, the scalar square of a vector is equal to the product of its coordinates, it follows that  $(\mathbf{a}^2) = a_1 a_2 < 0$  and  $(\mathbf{b}^2) = b_1 b_2 < 0$ . From this, it follows in particular that all numbers  $a_1, a_2, b_1, b_2$  are nonzero.

We obtain from formula (12.11) that  $\mathcal{U}(\mathbf{a}) = \alpha a_1 \mathbf{f}_1 + \alpha^{-1} a_2 \mathbf{f}_2$ , and the condition  $\langle \mathcal{U}(\mathbf{a}) \rangle = \langle \mathbf{b} \rangle$  means that  $\mathcal{U}(\mathbf{a}) = \mu \mathbf{b}$  for some  $\mu \neq 0$ . This yields the relationships  $\alpha a_1 = \mu b_1$  and  $\alpha^{-1} a_2 = \mu b_2$ , that is,

$$\mu = \frac{\alpha a_1}{b_1}, \quad a_2 = \alpha \mu b_2 = \frac{\alpha^2 a_1 b_2}{b_1}, \quad \alpha^2 = \frac{a_2 b_1}{a_1 b_2} = \frac{a_1 a_2 b_1 b_2}{(a_1 b_2)^2}.$$

It is obvious that the latter relationship can be solved for a real number  $\alpha$  if  $a_1 a_2 b_1 b_2 > 0$ , and this inequality is satisfied, since by assumption,  $a_1 a_2 < 0$  and  $b_1 b_2 < 0$ .  $\square$

Let us note that we have thus far not used motions of type (a). We shall need them to strengthen the theorem we have just proved. To do so, we shall make use of the notion of a flag, analogous to that introduced in Sect. 3.2 for real vector spaces.

**Definition 12.6** A *flag* in a space  $\mathbb{L}$  is a sequence of subspaces

$$\mathbb{L}_0 \subset \mathbb{L}_1 \subset \cdots \subset \mathbb{L}_n = \mathbb{L} \quad (12.17)$$

such that:

- (a)  $\dim \mathbb{L}_i = i$  for all  $i = 0, 1, \dots, n$ ;
- (b) each pair of subspaces  $(\mathbb{L}_{i+1}, \mathbb{L}_i)$  is directed.

A subspace  $\mathbb{L}_i$  is a hyperplane in  $\mathbb{L}_{i+1}$ , and as we have seen (see formula (12.5)), it defines a partition  $\mathbb{L}_{i+1}$  into two half-spaces:  $\mathbb{L}_{i+1} \setminus \mathbb{L}_i = \mathbb{L}_{i+1}^+ \cup \mathbb{L}_{i+1}^-$ . And as

earlier, the pair  $(\mathbb{L}_{i+1}, \mathbb{L}_i)$  is said to be *directed* if the order of the half-spaces is indicated, for example by denoting them by  $\mathbb{L}_{i+1}^+$  and  $\mathbb{L}_{i+1}^-$ . Let us note that in a flag defined by the sequence (12.17), the subspace  $\mathbb{L}_0$  has dimension 0, that is, it consists of a single point. We shall call this point the *center* of the flag (12.17).

**Theorem 12.7** *For any two flags of a hyperbolic space, there exists a motion taking the first flag to the second. Such a motion is unique.*

*Proof* In the space  $\mathbb{L}$ , let us consider two flags  $\Phi$  and  $\Phi'$  with centers at the points  $P \in \mathbb{L}$  and  $P' \in \mathbb{L}$ , respectively. Let  $O \in \mathbb{L}$  be the point corresponding to the line  $\langle e_{n+1} \rangle$  in  $\mathbb{L}$ , that is, the point with coordinates  $y_1 = 0, \dots, y_n = 0$  in relationship (12.4). By Theorem 12.5, there exist motions  $f$  and  $f'$  taking  $P$  to  $O$  and  $P'$  to  $O$ . Then the flags  $f(\Phi)$  and  $f'(\Phi')$  have their centers at the point  $O$ . Each flag is by definition a sequence of subspaces (12.17) in  $\mathbb{L}$  to which correspond the subspaces of the vector space  $\mathbb{L}$ . Thus to the flags  $f(\Phi)$  and  $f'(\Phi')$  there correspond two sequences of vector subspaces,

$$\langle e_{n+1} \rangle = \mathbb{L}_0 \subset \mathbb{L}_1 \subset \dots \subset \mathbb{L}_n = \mathbb{L} \quad \text{and} \quad \langle e_{n+1} \rangle = \mathbb{L}'_0 \subset \mathbb{L}'_1 \subset \dots \subset \mathbb{L}'_n = \mathbb{L},$$

where  $\dim \mathbb{L}_i = \dim \mathbb{L}'_i = i + 1$  for all  $i = 0, 1, \dots, n$ .

Let us recall that the space  $\mathbb{L}$  is identified with a part of the affine Euclidean space  $E$ , namely with the interior of the unit sphere  $U \subset E$  given by relationship (12.4). To investigate  $\mathbb{L}$  as a part of  $E$  (see Fig. 12.1), it will be convenient for us to associate with each subspace  $M \subset \mathbb{L}$  containing the vector  $e_{n+1}$ , the affine subspace  $N \subset E$  of dimension one less containing the point  $O$ . To this end, let us first associate with each subspace  $M \subset \mathbb{L}$  containing the vector  $e_{n+1}$ , the vector subspace  $N \subset M$  determined by the decomposition  $M = \langle e_{n+1} \rangle \oplus N$ . Employing notation introduced earlier, we obtain that

$$N = (\langle e_{n+1} \rangle^\perp \cap M) = (\langle e_1, \dots, e_n \rangle \cap M) \subset \langle e_1, \dots, e_n \rangle = E_0,$$

that is,  $N$  is contained in the space of vectors of the affine space  $E$ . Consequently, the vector subspace  $N \subset E_0$  determines a set of parallel affine subspaces in  $E$  that are characterized by their spaces of vectors coinciding with  $N$ . Such affine subspaces can be mapped to each other by a translation (see p. 296), and to determine one of them uniquely, it suffices simply to designate a point contained in this subspace. As such a point, we shall choose  $O$ . Then the vector subspace  $N \subset E_0$  uniquely determines the affine subspace  $N \subset E$ , where clearly,  $\dim N = \dim N = \dim M - 1$ .

Thus we have established a bijection between  $k$ -dimensional vector subspaces  $M \subset \mathbb{L}$  containing the vector  $e_{n+1}$  and  $(k - 1)$ -dimensional affine subspaces  $N \subset E$  containing the point  $O$ . Here clearly, the notions of directedness for the pair  $M' \subset M$  and  $N' \subset N$  coincide. In particular, flags  $f(\Phi)$  and  $f'(\Phi')$  of the space  $\mathbb{L}$  with center  $O$  correspond to two particular flags of the affine Euclidean space  $E$  with center at the point  $O$ .

By Theorem 8.40 (p. 316), in an affine Euclidean space, there exists for every pair of flags, a motion that takes the first flag to the second. Since in our case, both flags have a common center  $O$ , it follows that this motion has the fixed point  $O$ , and by Theorem 8.39, it is an orthogonal transformation  $\mathcal{A}$  of the Euclidean space  $E_0$ . Let us consider  $g = \mathbb{P}(\mathcal{A})$ , the motion of type (a) of the space  $\mathbb{L}$  corresponding to this orthogonal transformation  $\mathcal{A}$ . Clearly, it takes the flag  $f(\Phi)$  to  $f'(\Phi')$ , that is,  $gf(\Phi) = f'(\Phi')$ . From this, we obtain that  $f'^{-1}gf(\Phi) = \Phi'$ , as asserted in the theorem.

It remains to prove the assertion about uniqueness in the statement of the theorem. Let  $f_1$  and  $f_2$  be two motions taking some flag  $\Phi$  with center at the point  $P$  to the same flag, that is, such that  $f_1(\Phi) = f_2(\Phi)$ . Then  $f = f_1^{-1}f_2$  is a motion, and  $f(\Phi) = \Phi$ . If we prove that  $f$  is the identity transformation, then the required equality  $f_1 = f_2$  will follow.

By Theorem 12.5, there exists a motion  $g$  taking the point  $P$  to  $O$ . Let us set  $\Phi' = g(\Phi)$ . Then  $\Phi'$  is a flag with center at the point  $O$ . From the equalities  $f(\Phi) = \Phi$  and  $g(\Phi) = \Phi'$  it follows that  $gfg^{-1}(\Phi') = \Phi'$ . Let us denote the motion  $gfg^{-1}$  by  $h$ . It clearly takes the flag  $\Phi'$  to itself, and in particular, has the property that  $h(O) = O$ . From what we said on p. 438, it follows that  $h$  is a motion of type (a), that is,  $h = \mathbb{P}(\mathcal{A})$ , where  $\mathcal{A}$  is a Lorentz transformation of the space  $L$  that in turn, is determined by a certain orthogonal transformation  $\mathcal{U}$  of the Euclidean space  $E_0$ .

Let  $\Phi''$  be the flag in the Euclidean space  $E_0$  corresponding to the flag  $\Phi'$  of the space  $\mathbb{L}$ . Then from the condition  $h(\Phi') = \Phi'$ , it follows that  $\mathcal{U}(\Phi'') = \Phi''$ , that is, the transformation  $\mathcal{U}$  takes the flag  $\Phi''$  to itself. Consequently (see p. 225), the transformation  $\mathcal{U}$  is the identity, which yields that the motion  $h$  that it defines is the identity. From the relationship  $h = gfg^{-1}$ , it then follows that  $gf = g$ , that is,  $f$  is the identity transformation.  $\square$

Thus motions of a hyperbolic space possess the same property as that established in Sect. 8.4 (p. 317) for motions of affine Euclidean spaces. It is this that explains the special place of hyperbolic spaces in geometry. The Norwegian mathematician Sophus Lie called this property “free mobility.” There exists a theorem (which we shall not only not prove, but not even formulate precisely) showing that other than the space of Euclid and the hyperbolic space of Lobachevsky, there is only one space that exhibits this property, called a Riemann space (we shall have a bit to say about this in Sect. 12.3). This assertion is called the *Helmholtz–Lie theorem*. For its formulation, it would be necessary first of all to define just what we mean here by “space,” but we are not going to delve into this.

The property that we have deduced (Theorem 12.7) suffices for discussing the axiomatic foundations of hyperbolic geometry.

## 12.2 The Axioms of Plane Geometry\*

Hyperbolic geometry arose historically as a result of the analysis of the axiomatic systems of Euclidean geometry. The viewpoint toward geometry as based on a small



number of postulates from which all the remaining results are derived by way of formal proof arose in ancient Greece approximately in the sixth century B.C.E. Tradition connects this viewpoint with the name Pythagoras. An account of geometry with this point of view is contained in Euclid's *Elements* (third century B.C.E.). This point of view was accepted during the development of science in the modern era, and for a long time, geometry was taught directly from Euclid's books, and then later, there appeared simplified accounts. Moreover, this same point of view came to permeate all of mathematics and physics. In this spirit were written, for example, Newton's *The Mathematical Principles of Natural Philosophy*, known as the *Principia*. In physics and generally in the natural sciences, "laws of nature" played the role of axioms.

In mathematics, this direction of thought led to a more thorough working out of the axiom system of Euclidean geometry. Euclid divides the assertions on which his exposition is based into three types. One he calls "definitions"; another, "axioms"; and the third, "postulates" (the principle separating the last two of these is unclear to modern researchers). Many of his "definitions" also seem questionable. For example, the following: "A line is a length without width" (definitions of "length" and "width" are not given). Some "axioms" and "postulates" (we shall call all of these axioms) are simple corollaries of others, so that they could as well have been discarded. But what attracted the most attention was the "fifth postulate," which in Euclid is formulated thus:

That if a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.

This axiom differs from the others in that its formulation is notably more complex. Therefore, the following question arose (probably already in antiquity): can this assertion be proved as a theorem derived from the other axioms? An enormous number of "proofs of the fifth postulate" appeared, in which, however, there was always found a logical error. These investigations nevertheless helped in clarifying the situation. For example, it was proved that in the context of the other axioms, the fifth postulate is equivalent to the following assertion about parallel lines that is now usually presented as this postulate: through every point  $A$  not lying on a line  $a$ , it is possible to construct exactly one line  $b$  parallel to  $a$  (lines  $a$  and  $b$  are said to be parallel if they do not intersect). Here the *existence* of a line  $b$  parallel to  $a$  and passing through the point  $A$  can easily be proved. The entire content of the fifth postulate is reduced to the assertion about its *uniqueness*.

Finally, at the beginning of the nineteenth century, a number of researchers, one of whom was Nikolai Ivanovich Lobachevsky (1792–1856), came up with the idea that a proof of the fifth postulate is impossible, and so its *negation* leads to a new geometry, logically no less perfect than the geometry of Euclid, even though it contains in some respects some unusual propositions and relationships.

The question could be posed more precisely as a result of the development of the axiomatic method. This was done by Moritz Pasch (1843–1930), Giuseppe Peano (1858–1932), and David Hilbert (1862–1943) at the end of the nineteenth century. In his work on the foundations of geometry, Hilbert formulated in particular the

principles on which an axiomatic system is constructed. Today, such an approach has become commonplace; we used it to define vectors and Euclidean spaces. The general principle consists in fixing a certain set of *objects*, which remain undefined (for example, in the case of the definition of a vector space, these were scalars and vectors), and also in fixing certain *relations* that are to exist among these objects, which are likewise undefined (in the case of the definition of a vector space, these were addition of vectors and multiplication of a vector by a scalar). Finally, axioms are introduced that establish the specific properties of the introduced concepts (in the case of the definition of a vector space, these were enumerated in Sect. 3.1). With such a formulation, there remains only the question of *consistency* of the theory, that is, whether it is possible from the given axioms to derive simultaneously some statement as well as its negation. In the sequel, we shall introduce an axiom system for hyperbolic geometry (restriction to the case of dimension 2) and discuss the question of its consistency.

Let us begin with a discussion of axioms. The lists of axioms that Hilbert and his predecessors introduced in their early work turned out to possess certain logical defects. For example, in deduction, it turned out to be necessary to use certain assertions that were not contained among the axioms. Hilbert then supplemented his system of axioms. Later, this system of axioms was simplified for the sake of clarity. We shall use the axiom system proposed by the German geometer Friedrich Schur (1856–1932).<sup>1</sup> Here we shall restrict our attention (exclusively for the sake of brevity) to the axiomatics of the plane.

A *plane* is a certain set  $\Pi$ , whose elements  $A, B$ , and so on, are called *points*. Certain bijective mappings  $f : \Pi \rightarrow \Pi$  are called *motions*. These are the fundamental objects. The *relationships* among them are expressed as follows:

- (A) Certain distinguished subsets  $l, l'$ , and so on, of the set  $\Pi$  are called *lines*. That an element  $A \in \Pi$  belongs to the subset  $l$  is expressed by saying that “the point  $A$  lies on the line  $l$ ” or “the line  $l$  passes through the point  $A$ .”
- (B) For three given points  $A, B, C$  lying on a given line  $l$ , it is specified when the point  $C$  is considered to *lie between* the points  $A$  and  $B$ . This must be specified for every line  $l$  and for every three points lying on it.

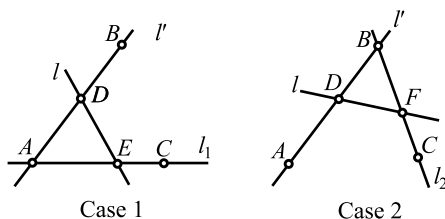
These objects and relations satisfy the conditions called *axioms*, which it is convenient to collect into several groups:

- I. Axioms of relationship
  - 1. For every two points, there exists a line passing through them.
  - 2. If these points are distinct, then such a line is unique.
  - 3. On every line there lie at least two points.
  - 4. For every line, there exists a point not lying on it.
- II. Axioms of order
  - 1. If on some line  $l$ , the point  $C$  lies between points  $A$  and  $B$ , then it is distinct from them and also lies between points  $B$  and  $A$ .

---

<sup>1</sup>Here we shall follow the ideas of Boris Nikolaevich Delaunay, or Delone (1890–1980), in his pamphlet *Elementary Proof of the Consistency of Hyperbolic Geometry*, 1956.

**Fig. 12.3** Intersection of the sides of a triangle by a line



2. If  $A$  and  $C$  are two distinct points on some line, then on this line there is at least one point  $B$  such that  $C$  lies between points  $A$  and  $B$ .
3. Among three points  $A$ ,  $B$ , and  $C$  lying on a given line, not more than one of the points lies between the two others.

Before formulating the last axiom of this group, let us give some new definitions. The set of all points  $C$  on a given line  $l$  passing through the points  $A$  and  $B$  that lie between them (including the points  $A$  and  $B$  themselves) is called a *segment* with endpoints  $A$  and  $B$ , and is denoted by  $[A, B]$ . Axiom 2 of group II can be reformulated thus:  $[A, C] \neq l \setminus (A \cup C)$ , with the inequality here being understood as an inequality of sets. That a segment  $[A, B]$  contains points other than  $A$  and  $B$  is proved on the basis of the axioms of group I and the last axiom of group II, to the formulation of which we now turn. Three points  $A, B, C$  not all lying on any one line are called a *triangle*, and this relationship is denoted by  $[A, B, C]$ . The segments  $[A, B]$ ,  $[B, C]$ , and  $[C, A]$  are called the *sides* of the triangle  $[A, B, C]$ .

4. Pasch's axiom. If points  $A, B, C$  do not all lie on the same line, none of them belong to the line  $l$ , and the line  $l$  intersects one side of the triangle  $[A, B, C]$ , then it also intersects another side of the triangle.

In other words, if a line  $l$  has a point  $D$  in common with the line  $l'$  passing through points  $A$  and  $B$ , with  $D$  lying between  $A$  and  $B$  on  $l'$ , then the line  $l$  either has a common point  $E$  with the line  $l_1$  passing through  $B$  and  $C$ , with  $E$  lying between them on  $l_1$ , or has a common point  $F$  with the line  $l_2$  passing through  $A$  and  $C$ , with  $F$  lying between them on  $l_2$ . The two cases discussed in this last axiom are depicted in Fig. 12.3.

### III. Axioms of motion

1. For every motion  $f$ , the inverse mapping  $f^{-1}$  (which exists by the definition of a motion as a bijective mapping of the set  $\Pi$ ) is also a motion.
2. The composition of two motions is a motion.
3. A motion preserves the order of points. That is, a motion  $f$  takes a line  $l$  to a line  $f(l)$ , and if the point  $C$  on the line  $l$  lies between points  $A$  and  $B$  on this line, then the point  $f(C)$  of the line  $f(l)$  lies between points  $f(A)$  and  $f(B)$ .

The formulation of the fourth axiom of motion requires certain results that can be obtained as corollaries of the axioms of relationship and order. We shall not prove these here, but let us give only the formulations.<sup>2</sup>

Let us begin with properties of lines. Let us choose a point  $O$  on a line  $l$ . Points  $A$  and  $B$  on this same line, both of them different from  $O$ , are said to lie *on one side* of  $O$  if  $O$  does not lie between  $A$  and  $B$ . If we select some point  $A$  different from  $O$ , then points  $B$  different from  $O$  and lying together with  $A$  on one side of  $O$  form a subset of the set of points of the line  $l$  called a *half-line* and denoted by  $l^+$ . It can be proved that if we choose in this subset another point  $A'$ , then the half-line formed with it will be the same as before. Here what is important is only the choice of the point  $O$ . If we choose a point  $A_1$  such that  $O$  lies between  $A$  and  $A_1$ , then the point  $A_1$  determines another half-line, denoted by  $l^-$ . The half-lines  $l^+$  and  $l^-$  determined by the points  $A$  and  $A_1$  do not intersect, and their union is  $l \setminus O$ , that is,  $l^+ \cap l^- = \emptyset$  and  $l^+ \cup l^- = l \setminus O$ .

One can verify analogous properties for a line  $l$  in the plane  $\Pi$ . Let us consider two points  $A$  and  $B$  that do not belong to the line  $l$ . One says that they lie *on one side* of  $l$  if either the line  $l'$  passing through them does not intersect the line  $l$ , or the lines  $l$  and  $l'$  intersect in a point  $C$  that does not lie between points  $A$  and  $B$  of the line  $l'$ . The set of points not lying on the line  $l$  and lying on the same side of  $l$  as the point  $A$  is called a *half-plane*. Again, it is possible to prove that with the choice of another point  $A'$  instead of  $A$  in this half-plane, we define the same set. There exist two points  $A$  and  $A'$  that do not belong to the same half-plane. However we select these points (given a fixed line  $l$ ), we will always obtain two subsets  $\Pi^+$  and  $\Pi^-$  of the plane  $\Pi$  such that  $\Pi^+ \cap \Pi^- = \emptyset$  and  $\Pi^+ \cup \Pi^- = \Pi \setminus l$ .

Suppose we are given a point  $O$  and a line  $l$  passing through it. If in the partition of  $l \setminus O$  into two half-lines, one of them is distinguished, and in the partition  $\Pi \setminus l$  into two half-planes, one of them is distinguished (for example, let us denote them by  $l^+$  and  $\Pi^+$ , respectively), then the pair  $(O, l)$  is called a *flag* and is denoted by  $\Phi$ . As follows from what was discussed in Sect. 12.1, this is a special case (for  $n = 2$ ) of the notion of a flag introduced earlier.

Every motion takes a flag to a flag, that is, if  $f$  is a motion and  $\Phi$  is the flag  $(O, l)$ , then the sets  $f(l)^+$  and  $f(l)^-$ , whose union is  $f(l) \setminus f(O)$ , coincide with  $f(l^+)$  and  $f(l^-)$ , where  $l^+$  and  $l^-$  are the half-lines on the line  $l$  determined by the point  $O$ . Here their order can change. Analogously, a pair of half-planes  $f(\Pi)^+$  and  $f(\Pi)^-$  defined by the line  $f(l)$  coincide with the pair  $f(\Pi^+)$  and  $f(\Pi^-)$ , where  $\Pi^+$  and  $\Pi^-$  are the half-planes determined by the line  $l$ . Their order also can change.

We can now formulate the last (fourth) axiom of motion:

4. Axiom of free mobility. For any two flags  $\Phi$  and  $\Phi'$ , there exists a motion  $f$  taking the first flag to the second, that is,  $f(\Phi) = \Phi'$ . Such a motion is unique, and it is uniquely determined by the flags  $\Phi$  and  $\Phi'$ .

<sup>2</sup>Some of these are proved in first courses in geometry, and in any case, elementary proofs of all of these results can be found in Chap. 2 of the book *Higher Geometry*, by N.V. Efimov (Mir, 1953).

## IV. Axiom of continuity

1. Let a set of points of some line  $l$  be represented arbitrarily as the union of two sets  $M_1$  and  $M_2$ , where no point of the set  $M_1$  lies between two points of the set  $M_2$ , and conversely. Then there exists a point  $O$  on the line  $l$  such that  $M_1$  and  $M_2$  coincide with the half-lines of  $l$  determined by the point  $O$ , to either of which the point  $O$  can be joined.

This axiom is also called *Dedekind's axiom*.

Axioms I–IV that we have presented are called axioms of “absolute geometry.” They hold for both Euclidean and hyperbolic geometry. These two geometries are distinguished by the addition of one axiom that deals with parallel lines. Let us recall that parallel lines are lines having no points in common. Thus in both cases, one more axiom is added:

## V. Axiom of parallel lines

1. In Euclidean geometry: For every line  $l$  and every point  $A$  not lying on it, there exists at most one line  $l'$  passing through the point  $A$  and parallel to  $l$ .
- 1'. In hyperbolic geometry: For every line  $l$  and every point  $A$  not lying on it, there exist at least two distinct lines  $l'$  and  $l''$  parallel to  $l$ .

The justified interest in precisely these two axioms is due to the fact that already in absolute geometry (that is, with only the axioms from groups I–IV), it is possible to prove that for every line  $l$  and every point  $A$  not on  $l$ , there exists at least one line  $l'$  passing through  $A$  and parallel to  $l$ .

It is now possible to formulate more precisely the goal that mathematics set for itself in the attempt to “prove the fifth postulate,” that is, to derive assertion 1 in group V of axioms from axioms in groups I–IV. But Lobachevsky (and other researchers of the same epoch) came to the conclusion that this was impossible, and this meant that the system comprising groups I–IV and axiom 1' was consistent.

Strictly speaking, we could have posed such questions even earlier, in connection with any of the theories that we encountered based on some system of axioms, such as the theory of vector spaces or that of Euclidean spaces. The question of the consistency of the concepts of vector spaces or Euclidean spaces is easily answered: it suffices to show (in the case of real spaces) examples of vector spaces over  $\mathbb{R}^n$  of any finite dimension or Euclidean spaces with inner product  $(\mathbf{x}, \mathbf{y}) = x_1 y_1 + \cdots + x_n y_n$ . Of course, this assumes the construction and proof of the consistency of the theory of the real numbers, but that lies outside the scope of our investigation, and we shall not consider it here. However, assuming as given that the properties of real numbers are defined and do not raise any doubts, we may, for example, say that if the system of axioms of a real vector space given in Sect. 3.1 were inconsistent, then we would be able to derive two mutually contradictory assertions about the space  $\mathbb{R}^n$ . However, any assertion about the space  $\mathbb{R}^n$  can be reduced by definition to an assertion about the real numbers, and then we would obtain a contradiction in the domain of real numbers.

The same question could be posed in relationship to Euclidean geometry, that is, with respect to the system of axioms consisting of axioms of groups I–IV and

axiom 1 of group V. Here the answer is in fact already known, since we have constructed the theory of affine Euclidean spaces (even in arbitrary dimension  $n$ ). It is easily ascertained that for  $n = 2$ , all the axioms of Euclidean geometry that we introduced are satisfied. Some refinements are perhaps necessary only in connection with the axioms of order.

These axioms do not require an inner product on the space and are formulated for an arbitrary real affine space  $V$  in Sect. 8.2. All the assertions constituting the axioms of order now follow directly from the properties of order of the real numbers, except only Pasch's axiom. Its idea is that if a line "enters" a triangle, then it must "exit" from it. Intuitively, this is quite convincing, but with our approach, we must derive this assertion from the properties of affine spaces. It is a very simple argument, whose details we leave to the reader.

Specifically, by what is given, points  $A$  and  $B$  (we shall use the same notation as in the formulation of the axioms) lie in different half-planes into which the line  $l$  divides the plane  $\Pi$ . Everything depends on the half-plane to which the point  $C$  belongs: to the same one as  $A$ , or to the same one as  $B$ . In the first case, the line  $l$  has a common point with the line  $l_2$ , which lies on it between  $B$  and  $C$ , while in the second case, the common point is with the line  $l_1$ , which lies between  $A$  and  $C$ ; see Fig. 12.3. In each of these two cases, the assertion of Pasch's axiom is easy to verify if we recall the definitions.

We in fact checked in one form or another that the remaining axioms are satisfied even as assertions that relate to arbitrary dimension.

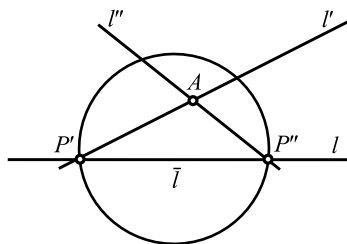
We shall now turn to the axioms of hyperbolic geometry, that is, the axioms of groups I–IV and axiom 1' of group V. We shall prove that they are consistent, based on the consistency of the usual properties (which likewise are easily reduced to certain axioms) of the set of real numbers  $\mathbb{R}$  and based on the theory of Euclidean spaces of dimension 2 and 3 constructed on this basis. On this foundation, we shall prove the following result.

**Theorem 12.8** *The system of axioms of hyperbolic geometry is consistent.*

*Proof* We shall consider in the Euclidean plane  $L$  the open disk  $K$  (given, for example, in some coordinate system by the condition  $x^2 + y^2 < 1$ ). We shall call the set of its points a "plane" (denoted by  $\overline{\Pi}$ ), and we shall call "points" only the points of this disk. The intersection of every line  $l$  of the plane  $L$  with the disk  $K$  that has at least one point in common with this disk is the interior of some segment (this was proved in the previous section). We shall call such nonempty intersections  $l \cap K$  "lines," denoted by  $\bar{l}, \bar{l}'$ , and so on. Finally, we shall call a projective transformation of the plane  $L$  taking the disk  $K$  into itself a "motion."

Since the definition of projective transformation assumes a study of the projective plane, and a projective space of dimension  $n$  and its projective transformations were defined in Chap. 9 in terms of a vector space of dimension  $n + 1$ , it follows that for the analysis of the hyperbolic plane, we must use here a notion connected with a three-dimensional vector space. However, it would not be difficult to give a formulation appealing only to properties of the Euclidean plane.

**Fig. 12.4** “Lines” and “points” of the hyperbolic plane



Now let us define the fundamental relationships between “lines” and “points.” That a “line”  $\bar{l}$  passes through a “point”  $A \in \bar{l}$  will be understood to mean the condition that the line  $l$  passes through the point  $A$ . Thus an arbitrary “line”  $\bar{l}$  is the set of “points” that lie on it. Let “points”  $A, B, C$  lie on the “line”  $\bar{l}$ . We shall say that a “point”  $C$  lies between “points”  $A$  and  $B$  if such is the case for  $A, B$ , and  $C$  as points on the Euclidean line  $l$  that contains  $\bar{l}$  (this makes sense, since  $l$  is contained in Euclidean space).

It remains to verify that the notions and relationships presented satisfy the axioms of hyperbolic geometry, that is, the axioms of groups I–IV and axiom 1' of group V. The verification of this for the axioms of groups I, II, and IV is trivial, since the corresponding objects and relationships are defined exactly as in the surrounding Euclidean plane. For the axioms of group III (axioms of motion), the required properties were proved in the previous section (indeed, for the case of a space of arbitrary dimension  $n$ ). It remains only to consider axiom 1' of group V.

Let  $\bar{l}$  be the “line” associated with the line  $l$  in the Euclidean plane  $L$ . Then the line  $l$  intersects the boundary  $S$  of the disk  $K$  in two different points:  $P'$  and  $P''$ . Let  $A$  be a “point” of the “plane”  $\bar{l}$  (that is, a point of the disk  $K$ ) not lying on the line  $l$ . By the axioms of Euclidean geometry, through the points  $A$  and  $P'$  in the plane  $L$ , there passes some line  $l'$ . It determines the “line”  $\bar{l}' = l' \cap K$  of the “plane”  $\bar{l}$ . Similarly, the point  $P''$  determines the “line”  $\bar{l}'' = l'' \cap K$ ; see Fig. 12.4.

The lines  $l'$  and  $l''$  are distinct, since they pass through different points  $P'$  and  $P''$  of the plane  $L$ . Therefore, by the axioms of Euclidean geometry, they have no common points other than  $A$ . But the “lines”  $\bar{l}'$  and  $\bar{l}''$ , as nonempty segments of Euclidean lines excluding the endpoints, contain infinitely many points and in particular, the “points”  $B' \in \bar{l}'$  and  $B'' \in \bar{l}''$ , with  $B' \neq B''$ . This means that the “lines”  $\bar{l}'$  and  $\bar{l}''$  are distinct. On the other hand, in the sense of our definitions, both of them are parallel to the “line”  $\bar{l}$ , that is, they have no common “points” with it (points of the disk  $K$ ). For example, the line  $l'$  has with  $l$  the common point  $P'$  in the Euclidean plane  $L$ , which means that by the axioms of Euclidean geometry, they have no other common points, and in particular, no common points in the disk  $K$ .

We see that assertion 1' holds for every “line”  $\bar{l} \subset \bar{l}$  and every “point”  $A \notin \bar{l}$ . Let us now assume that from the axioms of hyperbolic geometry there could be derived an inconsistency (that is, some assertion and its negation). Then we could apply the same reasoning to the notions that earlier, with the proof of Theorem 12.8, we wrote in quotation marks: “point,” “plane,” “line,” and “motion.” Since they, as we have seen, satisfy all the axioms of hyperbolic geometry, we would again

arrive at a contradiction. But the notions “plane,” “line,” and “motion,” and also the relationship “lies between” for three points on a line were defined in terms of Euclidean geometry. Thus we would arrive at a contradiction to Euclidean geometry itself.  $\square$

Let us focus attention on this fine logical construction: we construct objects in some domain that satisfy a certain system of axioms, and thus we prove the consistency of this system if the consistency of the domain from which the necessary objects are taken has been accepted. Today, one says that a *model* of this axiom system has thereby been constructed in another domain. In particular, we earlier constructed a model of hyperbolic geometry in the theory of vector spaces. Only by constructing such a model was the question of the provability of the “fifth postulate” decided in mathematics.

In conclusion, it is of interest to dwell a bit on the history of this question. Independent of Lobachevsky, a number of researchers came to the conclusion that a negation of the “fifth postulate” leads to a meaningful and consistent branch of mathematics, a “new geometry,” eventually given the designation “non-Euclidean geometry.” There is no question here of priority. All the researchers clearly worked independently of one another (Gauss’s correspondence from the 1820s, Lobachevsky’s publication of 1829, and János Bolyai’s of 1832). Most of these who became known later were amateurs, not professional mathematicians. But there were some exceptions: outside of Lobachevsky, there was the greatest mathematician of that epoch—Gauss. The majority of such researchers known to us who clearly arrived at the same conclusions independently became known precisely because of their correspondence with Gauss, which was published along with other of Gauss’s papers after his death. It is clear from these publications that in his youth, Gauss had attempted to prove the fifth postulate, but later concluded that there existed a meaningful and consistent geometry that did not include this postulate. In his letters, Gauss discussed the similar views of his correspondents with great interest.

He clearly received the work of Lobachevsky with sympathetic understanding when it began to appear in translation, and on Gauss’s recommendation, Lobachevsky was elected a member of the Göttingen Academy of Sciences.

In one of Gauss’s diaries can be seen the name Nikolai Ivanovich Lobachevsky, written in Cyrillic letters:

Н И К О Л А Й   И В А Н О В И Ч   Л О Б А Ч Е В С К И Й

But it is surprising that Gauss himself, throughout his entire life, published not a line on this subject. Why was that? The usual explanation is that Gauss was afraid of not being understood. Indeed, in one letter in which he touched on the question of the “fifth postulate” and non-Euclidean geometry, he wrote, “since I fear the clamor of the Boeotians.” But it seems that this cannot be the full explanation of his mysterious silence. In his other works, Gauss did not fear being misunderstood



by his readers.<sup>3</sup> It is possible, however, that there is another explanation for Gauss's silence. He was one of the few who realized that however many interesting theorems of non-Euclidean geometry might be deduced, this would prove nothing definitively; there would always remain the theoretical possibility that future derivations would yield a contradictory assertion. And perhaps Gauss understood (or sensed) that at the time (first half of the nineteenth century), the mathematical concepts had not yet been developed to pose and solve this question rigorously.

Apparently, Lobachevsky was among the small number of mathematicians in addition to Gauss who understood this. For him, as with Gauss, there stood the question of "incomprehensibility." First of all, for Lobachevsky, there was the lack of comprehension among Russian mathematicians, especially analysts, who totally failed to accept his work. In any case, he constantly attempted to find a consistent foundation for his geometry. For example, he discovered its striking parallel with *spherical geometry* and expressed the idea that it was the "geometry of the sphere with imaginary radius." His geometry could indeed have been realized in the form of some other model if the very notion of model had been sufficiently developed at that time.

Beyond this (as noted by the French mathematician André Weil (1906–1998)), here we have the simplest case of *duality* between compact and noncompact symmetric spaces, discovered in the twentieth century by Élie Cartan.

Moreover, Lobachevsky proved that in three-dimensional hyperbolic space, there is a surface (called today a *horosphere*) such that if we consider only the set of its points and take as lines the curves of a specific type lying on it (called today *horocycles*), then all the axioms of Euclidean geometry are satisfied. From this it follows that if hyperbolic geometry is consistent, then Euclidean geometry is also consistent. Even if we accept the hypothesis that the "fifth postulate" does not hold, Euclidean geometry is still realized on the horosphere. Thus in principle, Lobachevsky came very close to the concept of a model. But he did not succeed in constructing a model of hyperbolic geometry in the framework of Euclidean geometry. Such a construction was not easily granted to mathematicians.

The following paragraph offers only a hint, and not a precise formulation, of the corresponding assertions.

First, in 1868, Eugenio Beltrami (1835–1899) constructed in three-dimensional Euclidean space a certain surface called a *pseudosphere* or *Beltrami surface*, whose Gaussian curvature (see the definition on p. 265) at every point is the same negative number. Hyperbolic geometry can be realized on the pseudosphere, where the role of lines is played by so-called *geodesic lines*.<sup>4</sup> However, here we are talking about only a piece of the pseudosphere and a piece of the hyperbolic plane. Here the posing of the question must be radically changed, since the majority of the axioms that we have given assume (as in, for example, Euclidean geometry) the possibility

---

<sup>3</sup>For example, his first published book, *Disquisitiones Arithmeticae*, was considered for a long time to be quite inaccessible.

<sup>4</sup>More about this can be found, for example, in the book *A Course of Differential Geometry and Topology*, by A. Mishchenko and A. Fomenko (Mir, 1988).

of continuing lines to infinity. The coincidence of two bounded pieces is understood in the sense of the coincidence of the measures of lengths and angles, about which, in the case of hyperbolic geometry, more will be said in the following section. Moreover, Hilbert later proved that the hyperbolic plane cannot in this sense be completely identified with any surface in three-dimensional space (much later it was proved that it is possible for some surface in five-dimensional space).

The model of hyperbolic geometry that we gave for the proof of Theorem 12.8 was constructed by Felix Klein (1849–1925) in 1870. The history of its appearance was also astounding. Formally speaking, this model was constructed in 1859 by the English mathematician Arthur Cayley (1821–1895). But he considered it only as a certain construction in projective geometry and apparently did not notice the connection with non-Euclidean geometry. In 1869, the young (twenty-year-old) Klein became acquainted with his work. He recalled that in 1870, he gave a talk on the work of Cayley at the seminar of the famous mathematician Weierstrass, and, as he writes, “I finished with a question whether there might exist a connection between the ideas of Cayley and Lobachevsky. I was given the answer that these two systems were conceptually widely separated.” As Klein puts it, “I allowed myself to be convinced by these objections and put aside this already mature idea.” However, in 1871, he returned to this idea, formulated it mathematically, and published it. But then his work was not understood by many. In particular, Cayley himself was convinced as long as he lived that there was some logical error involved. Only after several years were these ideas fully understood by mathematicians.

Of course, one can ask not only about the existence of Euclidean and hyperbolic geometries, but also about a number of different (in a certain sense) geometries. Here we shall formulate only the results that are relevant to the current discussion.<sup>5</sup>

First of all, we must give a precise sense to what we mean by “different” or “identical” geometries. This can be done with the help of the notion of *isomorphism* of geometries, which is analogous to the notion of isomorphism of vector spaces introduced earlier. Within the framework of a system of axioms used in this section, this can be done as follows. Let  $\Pi$  and  $\Pi'$  be two planes satisfying the axioms of groups I–IV, and let  $G$  and  $G'$  be sets of motions of the respective planes. Mappings  $\varphi : \Pi \rightarrow \Pi'$  and  $\psi : G \rightarrow G'$  define an isomorphism  $(\varphi, \psi)$  of these geometries if the following conditions are satisfied:

- (1) Both mappings  $\varphi$  and  $\psi$  are bijections.
- (2) The mapping  $\varphi$  takes every line  $l$  in the plane  $\Pi$  to some line  $\varphi(l)$  in the plane  $\Pi'$ .
- (3) The mapping  $\varphi$  preserves the relationship “lies between.” This means that if points  $A$ ,  $B$ , and  $C$  lie on the line  $l$ , with  $C$  lying between  $A$  and  $B$ , then the point  $\varphi(C)$  lies between  $\varphi(A)$  and  $\varphi(B)$  on the line  $\varphi(l)$ .
- (4) The mappings  $\varphi$  and  $\psi$  agree in the following sense: for every motion  $f \in G$ , its image  $\psi(f)$  is equal to  $\varphi f \varphi^{-1}$ . This means that for every point  $A \in \Pi$ , the equality  $(\psi(f))(\varphi(A)) = \varphi(f(A))$  holds.

---

<sup>5</sup>Their proofs are given in every course in higher geometry, for example, in the book *Higher Geometry*, by N.V. Efimov, mentioned earlier.

- (5) For every motion  $f \in G$ , the equality  $\psi(f^{-1}) = \psi(f)^{-1}$  holds, and for every pair of motions  $f_1, f_2 \in G$ , we have  $\psi(f_1 f_2) = \psi(f_1)\psi(f_2)$ .

Let us note that some of these conditions can be derived from the others, but for brevity, we shall not do this.

We shall consider geometries up to isomorphism as just described, that is, we shall consider two geometries the same if there exists an isomorphism between them. In particular, geometries with respective axioms 1 and 1' in group V are clearly not isomorphic to each other, that is, they are two different geometries. From this point of view, geometries (in the plane) satisfying axioms 1 and 1' are fundamentally different from each other. Namely, it has been proved that all geometries satisfying axiom 1 in group V are isomorphic.<sup>6</sup> But geometries that satisfy axiom 1' in group V are characterized up to isomorphism by a certain real number  $c$  called their *curvature*. This number is usually assumed to be negative, and then it can take on any value  $c < 0$ .

Klein suggested that Euclidean geometry can be viewed as the limiting case of hyperbolic geometry as the curvature  $c$  approaches zero.<sup>7</sup> As Klein further observed, if axiom 1 (of Euclid) is satisfied in our world, then we shall never know it. Since every physical measurement is taken with a certain degree of error, to establish the precise equality  $c = 0$  is impossible, for there always remains the possibility that the number  $c$  is less than zero, but it is so small in absolute value that it lies beyond the limits of our measurements.

## 12.3 Some Formulas of Hyperbolic Geometry\*

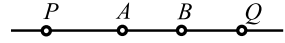
First of all, we shall define the distance between points in the hyperbolic plane using its definition as the set of points of the projective plane  $\mathbb{P}(L)$  corresponding to the lines of the three-dimensional pseudo-Euclidean space  $L$  lying within the light cone and its interpretation as the set of points on the unit circle  $U$  in the affine Euclidean plane  $E$ ; see Sect. 12.1.

The meaning of the notion of distance is that it should be preserved under motions of the hyperbolic plane. But we have defined a motion as a certain special projective transformation  $\mathbb{P}(\mathcal{A})$  of the projective plane  $\mathbb{P}(L)$ . Theorem 9.16 shows that in general, it is impossible to associate a number that does not change under an arbitrary projective transformation not only with two points, but even with three points of the projective line. But we shall use the fact that motions of the hyperbolic plane are not arbitrary projective transformations  $\mathbb{P}(L)$ , but only those that take the light cone in the space  $L$  into itself.

Namely, to two arbitrary points  $A$  and  $B$  correspond the lines  $\langle a \rangle$  and  $\langle b \rangle$ , lying inside the light cone. We shall show that they determine two additional points,  $P$

<sup>6</sup>Of course, here we are assuming that they all satisfy the axioms of groups I–IV.

<sup>7</sup>Felix Klein. *Nicht-Euklidische Geometrie*, Göttingen, 1893. Reprinted by AMS Chelsea, 2000.

**Fig. 12.5** The segment  $[PQ]$ 

and  $Q$ , that correspond to lines lying on the light cone. But *four points* of a projective space lying on a line already determine a number that does not change under arbitrary projective transformations, namely their cross ratio (defined in Sect. 9.3). We shall use this number for defining the distance between points  $A$  and  $B$ . This definition has the special feature that it uses points corresponding to lines lying on the light cone ( $P$  and  $Q$ ), which are thus not points of the hyperbolic plane.

We shall assume that the points  $A$  and  $B$  are distinct (if they coincide, then the distance between them is zero by definition). This means that the vectors  $\mathbf{a}$  and  $\mathbf{b}$  are linearly independent. It is obvious that then a unique projective line  $l$  passes through these points; it corresponds to the plane  $L' = \langle \mathbf{a}, \mathbf{b} \rangle$ . The line  $l$  determines a line  $l'$  in the affine Euclidean space  $E$ , depicted in Figs. 12.1 and 12.2. Since the line  $l'$  contains the points  $A$  and  $B$ , which lie inside the circle  $U$ , it intersects its boundary in *two* points, which we shall take as  $P$  and  $Q$ . This was in fact already proved in Sect. 12.1, but we shall now repeat the corresponding argument.

The points of  $l$  are the lines  $\langle \mathbf{x} \rangle$  consisting of all vectors proportional to the vectors  $\mathbf{x} = \overrightarrow{OA} + t\overrightarrow{AB}$ , where  $t$  is an arbitrary real number. Here the vector  $\overrightarrow{OA}$  equals  $\mathbf{a}$ , and the vector  $\overrightarrow{AB} = \mathbf{c}$  belongs to the subspace  $E_0$  if we assume that the points  $A, B$  and the line  $l$  lie in the affine space  $E$ . This means that  $\mathbf{x} = \mathbf{a} + t\mathbf{c}$ , where the vector  $\mathbf{c}$  can be taken as fixed, and the number  $t$  as variable. Points  $\mathbf{x}$  at the intersection of the line  $l'$  with the light cone  $V \subset L$  are given by the condition  $(\mathbf{x}^2) = 0$ , that is,

$$((\mathbf{a} + t\mathbf{c})^2) = (\mathbf{a}^2) + 2(\mathbf{a}, \mathbf{c})t + (\mathbf{c}^2)t^2 = 0. \quad (12.18)$$

We know that  $(\mathbf{a}^2) < 0$ , and the vector  $\mathbf{c}$  belongs to  $E_0$ . Since  $E_0$  is a Euclidean space and the points  $A$  and  $B$  are distinct, it follows that  $(\mathbf{c}^2) > 0$ . From this it follows that the quadratic equation (12.18) in the unknown  $t$  has two real roots  $t_1$  and  $t_2$  of opposite signs. Suppose for the sake of definiteness that  $t_1 < t_2$ . Then for  $t_1 < t < t_2$ , the value of  $((\mathbf{a} + t\mathbf{c})^2)$  is negative, and all points of the line  $l'$  corresponding to the values  $t$  in this interval belong to  $\mathbb{L}$ . We see that the line  $l$  intersects the light cone  $V$  in two points corresponding to the values  $t = t_1$  and  $t = t_2$ , while the values  $t_1 < t < t_2$  are associated with the points of the line  $\mathbb{L}_1$  (that is, one-dimensional hyperbolic space) passing through  $A$  and  $B$ . Thus the line  $\mathbb{L}_1$  coincides with the line segment  $l \subset E$  whose endpoints are  $P$  and  $Q$ , which correspond to the values  $t = t_1$  and  $t = t_2$ ; see Fig. 12.5.

It is clear that point  $A$  is contained in the interval  $(P, Q)$ . Applying the same argument to the point  $B$ , we obtain that the point  $B$  is also in the interval  $(P, Q)$ .

Let us label the points  $P$  and  $Q$  in such a way that  $P$  will denote the endpoint of the interval  $(P, Q)$  that is closer (in the sense of Euclidean distance) to the point  $A$ , and by  $Q$  the endpoint that is closer to  $B$ , as depicted in Fig. 12.5.

Now it is possible to give a definition of the distance between points  $A$  and  $B$ , which we shall denote by  $r(A, B)$ :

$$r(A, B) = \log DV(A, B, Q, P), \quad (12.19)$$

where  $DV(A, B, Q, P)$  is the cross ratio (see p. 337). Let us note that in the definition (12.19), we have not indicated the base of the logarithm. We could take any base greater than 1, since a change in base results simply in multiplying all distances by some fixed positive constant. But in any case, the length of a segment  $AB$  can be defined only up to a multiplicative factor that corresponds to the arbitrariness in the selection of a unit length on a line.

We shall explain a bit later why the logarithm appears in definition (12.19). The reason for using the cross ratio is explained by the following theorem.

**Theorem 12.9** *The distance  $r(A, B)$  does not change under any motion  $f$  of the hyperbolic plane, that is,  $r(f(A), f(B)) = r(A, B)$ .*

*Proof* The assertion of the theorem follows at once from the fact that a motion  $f$  of the hyperbolic plane is determined by a certain projective transformation  $\mathbb{P}(\mathcal{A})$ . This transformation  $\mathbb{P}(\mathcal{A})$  carries the line  $l'$  passing through points  $A$  and  $B$  to the line passing through the points  $\mathbb{P}(\mathcal{A})(A)$  and  $\mathbb{P}(\mathcal{A})(B)$ . This means that the transformation takes the points  $P$  and  $Q$ , the intersection of the line  $l'$  with the boundary of the disk  $U$ , to the points  $P'$  and  $Q'$ , the intersection of the line  $\mathbb{P}(\mathcal{A})(l')$  with this boundary. That is,  $P' = \mathbb{P}(\mathcal{A})(P)$  and  $Q' = \mathbb{P}(\mathcal{A})(Q)$ , or conversely,  $Q' = \mathbb{P}(\mathcal{A})(P)$  and  $P' = \mathbb{P}(\mathcal{A})(Q)$ . Moreover, the transformation  $\mathbb{P}(\mathcal{A})$  preserves the cross ratio of four points on a line (Theorem 9.17).  $\square$

To explain the role of the cross ratio, we jumped a bit ahead and skipped the verification that the argument of the logarithm in formula (12.19) was a number greater than 1 and also that in the definition of  $r(A, B)$ , all the conditions entering into the definition of a distance (p. xvii) were satisfied. We now return to this.

Let us assume that the points  $P, A, B, Q$  are arranged in the order shown in Fig. 12.5. For the cross product, we may use formula (9.28),

$$DV(A, B, Q, P) = \frac{|AQ| \cdot |PB|}{|BQ| \cdot |PA|} > 1, \quad (12.20)$$

since clearly,  $|AQ| > |BQ|$  and  $|PB| > |PA|$ . Therefore, the argument of the logarithm in formula (12.19) is a number greater than 1, and so the logarithm is a positive real number. Therefore,  $r(A, B) > 0$  for all pairs of distinct points  $A$  and  $B$ .

Let us note that it would be possible to make do without the order of the points  $P$  and  $Q$  that we chose. For this, it would be sufficient to verify (this follows directly from the definition of the cross ratio) that under a transposition of the points  $P$  and  $Q$ , the cross ratio  $d$  is converted into  $1/d$ . Thus the logarithm (12.19) that gives the distance is defined up to sign, and we can define the distance as the absolute value.

If we interchange the positions of  $A$  and  $B$ , then the points  $P$  and  $Q$  defined in the agreed-upon way also exchange places. It is easy to verify that the cross ratio determines a distance according to formula (12.19) that will not change. In other words, we have the equality

$$r(B, A) = r(A, B). \quad (12.21)$$

For any third point  $C$  collinear with  $A$  and  $B$  and lying between them, the condition

$$r(A, B) = r(A, C) + r(C, B) \quad (12.22)$$

is satisfied. It follows from the fact that (in the notation we have adopted)

$$DV(A, B, Q, P) = \frac{|AQ| \cdot |BP|}{|BQ| \cdot |AP|} = DV(A, C, Q, P) \cdot DV(C, B, Q, P), \quad (12.23)$$

since

$$DV(A, C, Q, P) = \frac{|AQ| \cdot |CP|}{|CQ| \cdot |AP|}, \quad DV(C, B, Q, P) = \frac{|CQ| \cdot |BP|}{|BQ| \cdot |CP|}. \quad (12.24)$$

For the verification, it remains only to substitute the expressions (12.24) into formula (12.23).

In any sufficiently complete course in geometry, it is proved without using the parallel postulate (that is, in the framework of “absolute geometry”) that there exists a function  $r(A, B)$  of a pair of points  $A$  and  $B$  that satisfies the following conditions:

1.  $r(A, B) > 0$  if  $A \neq B$ , and  $r(A, B) = 0$  if  $A = B$ ;
2.  $r(B, A) = r(A, B)$  for all points  $A$  and  $B$ ;
3.  $r(A, B) = r(A, C) + r(C, B)$  for every point  $C$  collinear with  $A$  and  $B$  and lying between them;

and most importantly,

4. the function  $r(A, B)$  is invariant under motions.

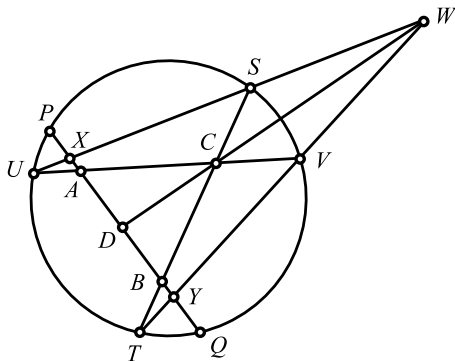
Using the definitions given at the beginning of this book, we may say in short that  $r(A, B)$  is a metric on the set of points in the plane under consideration and motions are isometries of this metric space.

Such a function is unique if we fix two distinct points  $A_0$  and  $B_0$  for which  $r(A_0, B_0) = 1$  (“unit of measurement”). This means that these assertions also hold in hyperbolic geometry, and formula (12.19) defines this distance (and the base of the logarithm in (12.19) is chosen in correspondence with the chosen “unit of measurement”).

Every triple of points  $A, B, C$  satisfies the condition

$$r(A, B) \leq r(A, C) + r(B, C). \quad (12.25)$$

**Fig. 12.6** The triangle inequality



This is the familiar *triangle inequality*, and in many courses in geometry, it is derived without use of the parallel postulate, that is, as a theorem of “absolute geometry.” Thus inequality (12.25) holds as well in hyperbolic geometry. But we shall now give a direct (that is, resting directly on formula (12.19)) proof of this due to Hilbert.

Let us recall that in the model that we have considered, the points of the hyperbolic plane are points of the disk  $K$  in the Euclidean plane  $L$ , and the lines of the hyperbolic plane are the line segments of the plane  $L$  that lie inside the disk  $K$ .

Let us consider three points  $A, B, C$  in the disk  $K$ . We shall denote the points of intersection of a line passing through  $A$  and  $B$  with the boundary of the disk  $K$  by  $P$  and  $Q$ , and the analogous points for the line passing through  $A$  and  $C$  will be denoted by  $U$  and  $V$ , and for the line passing through  $B$  and  $C$ , by  $S$  and  $T$ . See Fig. 12.6.

Let us denote the point of intersection of the line  $AB$  and the line  $SU$  by  $X$ , and the point of intersection of the line  $AB$  and the line  $TV$  by  $Y$ . Then we have the inequality

$$DV(A, B, Y, X) \geq DV(A, B, Q, P). \quad (12.26)$$

Indeed, the left-hand side of (12.26) is equal by definition to

$$DV(A, B, Y, X) = \frac{|AY| \cdot |BX|}{|BY| \cdot |AX|}, \quad (12.27)$$

and its right-hand side is given by the relationship (12.20). Therefore, inequality (12.26) follows from the fact that

$$\frac{|AY|}{|BY|} > \frac{|AQ|}{|BQ|} \quad \text{and} \quad \frac{|BX|}{|AX|} > \frac{|BP|}{|AP|}. \quad (12.28)$$

Let us prove the first of inequalities (12.28). Let us define  $a = |AB|$ ,  $t_1 = |BQ|$ , and  $t_2 = |BY|$ . Then we obviously obtain the expressions  $|AQ|/|BQ| = (a + t_1)/t_1$  and  $|AY|/|BY| = (a + t_2)/t_2$ . For  $a > 0$ , the function  $(a + t)/t$  in the variable  $t$  decreases monotonically with increasing  $t$ , and therefore, from the fact that  $t_2 < t_1$  (which is obvious from Fig. 12.6) follows the first of inequalities (12.28). Defining

$a = |AB|$ ,  $t_1 = |AX|$ , and  $t_2 = |AP|$ , using completely analogous arguments, we may prove the second inequality of (12.28).

Let us denote the intersection of the lines  $SU$  and  $by  $W$ , let us connect this line with the point  $C$ , and let us denote the point of intersection of the line thus obtained with the line  $AB$  by  $D$ . Then the points  $X, A, D, Y$  and points  $U, A, C, V$  are obtained from each other by a perspective mapping just as was done for the points  $Y, B, D, X$  and  $T, B, C, S$ . Then in view of Theorem 9.19, we have the relationships$

$$\frac{|AY| \cdot |DX|}{|DY| \cdot |AX|} = \frac{|AV| \cdot |CU|}{|CV| \cdot |AU|}, \quad \frac{|BX| \cdot |DY|}{|DX| \cdot |AY|} = \frac{|BS| \cdot |CT|}{|CS| \cdot |BT|}.$$

Multiplying these equalities, we have

$$\frac{|AY| \cdot |BX|}{|BY| \cdot |AX|} = \frac{|AV| \cdot |CU|}{|CV| \cdot |AU|} \cdot \frac{|BS| \cdot |CT|}{|CS| \cdot |BT|}.$$

Taking the logarithm of the last equality, and taking into account (12.27) for  $DV(A, B, Y, X)$ , the analogous expression for  $DV(A, C, U, V)$  and that for  $DV(B, C, S, T)$ , and definition (12.19), we obtain the relationship

$$\log DV(A, B, Y, X) = r(A, C) + r(B, C),$$

from which, taking into account (12.26), we obtain the required inequality (12.25).

Let us note that if the point  $B$  approaches  $Q$  along the segment  $PQ$  (see Fig. 12.6), then  $|BQ|$  approaches zero, and consequently,  $r(A, B)$  approaches infinity. This means that despite that fact that the line passing through the points  $A$  and  $B$  is represented in our figure by a segment of finite length, its length in the hyperbolic plane is infinite.

The measurement of angles is similar to that of line segments. As we know, an arbitrary point  $O$  on a line  $l$  partitions it into two half-lines. One half-line together with the point  $O$  is called a *ray*  $h$  with center  $O$ . Two rays  $h$  and  $k$  with common center  $O$  are called an *angle*; we shall assume that the ray  $h$  is obtained from  $k$  by a counterclockwise rotation. This angle is denoted by  $\angle(h, k)$ .

In "absolute geometry," it is proved that for each angle with vertex at the point  $O$ , there is a unique real number  $\angle(h, k)$  satisfying the following four conditions:

1.  $\angle(h, k) > 0$  for all  $h \neq k$ ;
2.  $\angle(k, h) = \angle(h, k)$ ;
3. if  $f$  is a motion and  $f(h) = h'$ ,  $f(k) = k'$ , and  $O' = f(O)$  is the vertex of the angle  $\angle(h', k')$ , then  $\angle(h', k') = \angle(h, k)$ .

To formulate the fourth property, we must introduce some additional concepts. Let the rays  $h$  and  $k$  forming the angle  $\angle(h, k)$  lie on lines  $l_1$  and  $l_2$ . The points in the plane lying on the same side of the line  $l_1$  as the points of the half-line  $k$  and on the same side of the line  $l_2$  as the points of the half-line  $h$  are called *interior points* of the angle  $\angle(h, k)$ . A ray  $l$  with the same center  $O$  as the rays  $h$  and  $k$  is said to be an *interior ray* of the angle  $\angle(h, k)$  if it consists of interior points of this angle.

We can now formulate the last property:



4. If  $l$  is an interior ray of the angle  $\angle(h, k)$ , then  $\angle(h, l) + \angle(l, k) = \angle(h, k)$ .

As in the case of distance between points, the measure of an angle is defined uniquely if we choose a “unit measurement,” that is, if we take a particular angle  $\angle(h_0, k_0)$  as the “unit angle measure.”

We shall point out an explicit method of defining the measure of angles in hyperbolic geometry that is realized in the disk  $K$  given by the relationship  $x^2 + y^2 < 1$  in the Euclidean plane  $L$  with coordinates  $x, y$ .

Let  $\angle(h', k')$  be the angle with center at the point  $O'$ , and let  $f$  be an arbitrary motion taking the point  $O'$  to the center  $O$  of the disk  $K$ . From the definitions, it is obvious that  $f$  takes the half-lines  $h'$  and  $k'$  to some half-lines  $h$  and  $k$  with center at the point  $O$ . Let us set the measure of  $\angle(h', k')$  equal to the Euclidean angle between the half-lines  $h$  and  $k$ . The main difficulty in this definition is that it uses a motion  $f$ , and therefore, we must prove that the measure of the angle thus obtained does not depend on the choice of the motion  $f$  (of course, with the condition  $f(O') = O$ ).

Let  $g$  be another motion with the same property that  $g(O') = O$ . Then  $g^{-1}(O) = O'$ , and this means that  $fg^{-1}(O) = O$ , that is, the motion  $fg^{-1}$  leaves the point  $O$  fixed. As we saw in Sect. 12.1 (p. 438), a motion possessing such a property is of type (a), which means that  $fg^{-1}$  corresponds to an orthogonal transformation of the Euclidean plane  $L$ ; that is, the angle  $\angle(\bar{h}, \bar{k})$  is taken to the angle  $\angle(h, k)$  via the orthogonal transformation  $fg^{-1}$ , which preserves the inner product in  $L$  and therefore does not change the measure of angles. This proves the correctness of the definition of angle measure that we have introduced. Equally easy are the verifications of properties 1–3.

The best-known property of angles in hyperbolic geometry is the following.

**Theorem 12.10** *In hyperbolic geometry, the sum of the angles of a triangle is less than two right angles, that is, less than  $\pi$ .*

Since we are talking about a triangle, we can restrict our attention to the plane in which this triangle lies and assume that we are working in the hyperbolic plane. The key result is related to the fact that an angle  $\angle(h, k)$  in hyperbolic geometry also determines a Euclidean angle, and we may then compare the measures of these angles. We shall denote the measure of the angle  $\angle(h, k)$  in hyperbolic geometry, as before, by  $\angle(h, k)$ , and its Euclidean measure by  $\angle_E(h, k)$ .

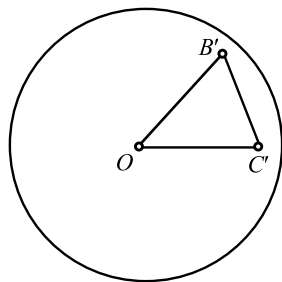
**Lemma 12.11** *If one ray of the angle  $\angle(h, k)$  (for example,  $h$ ) passes through the center  $O$  of the disk  $K$ , then the measure of this angle in the sense of hyperbolic geometry is less than the Euclidean measure, that is,*

$$\angle(h, k) < \angle_E(h, k). \quad (12.29)$$

First, we shall show how easily Theorem 12.10 follows from the lemma, and then we shall prove the lemma itself.

*Proof of Theorem 12.10* Let us denote the vertices of the triangle in question by  $A, B, C$ . Since the measure of an angle is invariant under a motion, it follows by

**Fig. 12.7** A triangle in the hyperbolic plane



Theorem 12.5 that we can choose a motion taking one of the vertices of the triangle (for example,  $A$ ) to the center  $O$  of the disk  $K$ . Let the vertices  $B$  and  $C$  be taken to  $B'$  and  $C'$ . See Fig. 12.7.

It suffices to prove the theorem for the triangle  $OB'C'$ . But for the angle  $\angle B'OC'$ , we have by definition the equality

$$\angle B'OC' = \angle_E B'OC',$$

and for the two remaining angles, we have by the lemma, the inequalities

$$\angle OB'C' < \angle_E OB'C', \quad \angle OC'B' < \angle_E OC'B'.$$

Adding, we obtain for the sum of the angles of triangle  $OB'C'$  the inequality

$$\angle B'OC' + \angle OB'C' + \angle OC'B' < \angle_E B'OC' + \angle_E OB'C' + \angle_E OC'B'.$$

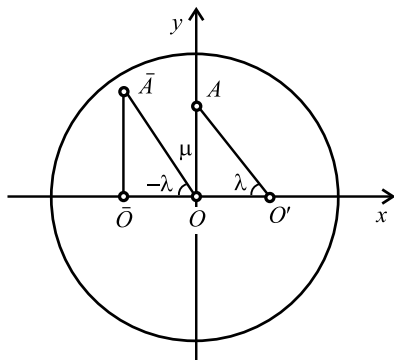
By a familiar theorem of Euclidean geometry, the sum on the right-hand side is equal to  $\pi$ , and this proves Theorem 12.10.  $\square$

*Proof of Lemma 12.11* We shall have to use the explicit form of the definition of the measure of an angle. Let the ray  $h$  of the angle  $\angle(h, k)$  pass through the point  $O$ . To describe the disk  $K$ , we shall introduce a Euclidean rectangular system of coordinates  $(x, y)$  and assume that the vertex of angle  $\angle(h, k)$  is located at the point  $O'$  with coordinates  $(\lambda, 0)$ , where  $\lambda \neq 0$ . For this, it is necessary to execute a rotation about the center of the disk in such a way that the point  $O'$  passes through some point of the line  $y = 0$  and use the fact that angles are invariant under such a rotation.

Now we must write down explicitly a motion  $f$  of the hyperbolic plane taking the point  $O$  to  $O'$ . We already constructed such a motion in Sect. 12.1; see Example 12.4 on p. 439. There, we proved that there exists a motion of the hyperbolic plane that takes the point with coordinates  $(x, y)$  to the point with coordinates  $(x', y')$ , given by the relationships

$$x' = \frac{ax + b}{bx + a}, \quad y' = \frac{y}{bx + a}, \quad a^2 - b^2 = 1. \quad (12.30)$$

**Fig. 12.8** Angles in the hyperbolic plane



If we want the point  $O' = (\lambda, 0)$  to be sent to the origin  $O = (0, 0)$ , then we should set  $a\lambda + b = 0$ , or equivalently,  $\lambda = -b/a$ . It is not difficult to verify that it is possible to represent any number  $\lambda$  in this form. Thus the mapping (12.30) has the form

$$x' = \frac{x - \lambda}{1 - \lambda x}, \quad y' = \frac{y}{a(1 - \lambda x)}. \quad (12.31)$$

Let the ray  $k$  intersect the  $y$ -axis at the point  $A$  with coordinates  $(0, \mu)$ ; see Fig. 12.8. (We note that this point is not required to be in the disk  $K$ .)

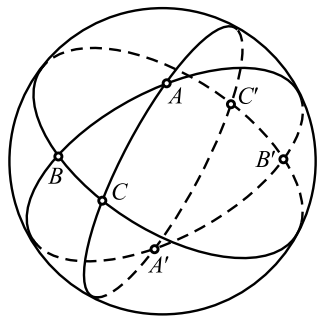
From formula (12.31), it is clear that our transformation takes a vertical line  $x = c$  to a vertical line  $x = c'$ . The point  $O$  is taken to the point  $\bar{O} = (-\lambda, 0)$ , the point  $A = (0, \mu)$  to the point  $\bar{A} = (-\lambda, \mu/a)$ , and the vertical line  $OA$  to the vertical line  $\bar{O}\bar{A}$ . By the definition of an angle in hyperbolic geometry,  $\angle O O' A = \angle_E \bar{O} \bar{O} \bar{A}$ . The tangents of the Euclidean angles are known to us:

$$\tan(\angle_E O O' A) = \frac{\mu}{\lambda}, \quad \tan(\angle_E \bar{O} \bar{O} \bar{A}) = \frac{\overline{OA}}{\bar{\lambda}} = \frac{\mu}{\lambda a};$$

see Fig. 12.8. Since  $a^2 = 1 + b^2$ , we have  $a > 1$ , and we see that in Euclidean geometry, we have the inequality  $\tan(\angle_E \bar{O} \bar{O} \bar{A}) < \tan(\angle_E O O' A)$ . The tangent is a strictly increasing function, and therefore we have the inequality  $\angle_E \bar{O} \bar{O} \bar{A} < \angle_E O O' A$  for angles that are Euclidean. But  $\angle O O' A = \angle_E \bar{O} \bar{O} \bar{A}$ , and this means that  $\angle O O' A < \angle_E O O' A$ .  $\square$

It is of interest to compare Theorem 12.10 with the analogous result for *spherical geometry*. We have not yet encountered spherical geometry in this course, even though it was developed in detail much earlier than hyperbolic geometry, indeed in antiquity. In spherical geometry, the role of lines is played by great circles on the sphere, that is, sections of the sphere obtained by all possible planes passing through its center. The analogy between great circles on the sphere and lines in the plane consists in the fact that the arc of the great circle joining points  $A$  and  $B$  has length no greater than that of any other curve on the sphere with endpoints  $A$  and  $B$ . This arc length of a great circle (which, of course, depends also on the radius  $R$  of the sphere) is called the *distance* on the sphere from point  $A$  to point  $B$ .

**Fig. 12.9** A triangle on the sphere



The measurement of lengths and angles on the sphere can generally be defined in exactly the same way as in Euclidean or hyperbolic geometry. Here the angle between two “lines” (that is, great circles) is equal to the value of the dihedral angle formed by the planes passing through these great circles. We have the following result.

**Theorem 12.12** *The sum of the angles of a triangle on the sphere is greater than two right angles, that is, greater than  $\pi$ .*

*Proof* Let there be given a triangle with vertices  $A, B, C$  on a sphere of radius  $R$ . Let us draw all the great circles whose arcs are the sides  $AB, AC$ , and  $BC$  of triangle  $ABC$ . See Fig. 12.9.

Let us denote by  $\Sigma_A$  the part of the sphere enclosed between the great circle passing through the points  $A, B$  and the great circle passing through  $A, C$ . We introduce the analogous notation  $\Sigma_B$  and  $\Sigma_C$ . Let us denote by  $\widehat{A}$  the measure of the dihedral angle  $\widehat{BAC}$  and similarly for  $\widehat{B}$  and  $\widehat{C}$ . Then the assertion of the theorem is equivalent to asserting that  $\widehat{A} + \widehat{B} + \widehat{C} > \pi$ .

But it is easy to see that the area of  $\Sigma_A$  is the same fraction of the area of the sphere as  $2\widehat{A}$  is of  $2\pi$ . Since the area of the sphere is equal to  $4\pi R^2$ , it follows that the area of  $\Sigma_A$  is equal to

$$4\pi R^2 \cdot \frac{2\widehat{A}}{2\pi} = 4R^2 \widehat{A}.$$

Similarly, we obtain expressions for the areas  $\Sigma_B$  and  $\Sigma_C$ ; they are equal to  $4R^2 \widehat{B}$  and  $4R^2 \widehat{C}$  respectively. Let us now observe that the regions  $\Sigma_A, \Sigma_B$ , and  $\Sigma_C$  together cover the entire sphere. Here each point of the sphere not part of triangle  $ABC$  or of triangle  $A'B'C'$  symmetric to it on the sphere belongs to only one of the regions  $\Sigma_A, \Sigma_B$ , and  $\Sigma_C$ , and every point in triangle  $ABC$  or the symmetric triangle  $A'B'C'$  is contained in all three regions. We therefore have

$$4R^2(\widehat{A} + \widehat{B} + \widehat{C}) = 4\pi R^2 + 2S_{\triangle ABC} + 2S_{\triangle A'B'C'} = 4\pi R^2 + 4S_{\triangle ABC}.$$

From this we obtain the relationship

$$\widehat{A} + \widehat{B} + \widehat{C} = \pi + \frac{S_{\triangle ABC}}{R^2}, \quad (12.32)$$

from which it follows that  $\widehat{A} + \widehat{B} + \widehat{C} > \pi$ . □

Formula (12.32) gives an example of a series of relationships systematically developed by Lobachevsky: if we were to assume that  $R^2 < 0$  (that is,  $R$  is a purely imaginary number), then clearly, we would obtain from (12.32) the inequality

$$\widehat{A} + \widehat{B} + \widehat{C} < \pi,$$

which is Theorem 12.10 of hyperbolic geometry. This is why Lobachevsky considered that his geometry is realized “on a sphere of imaginary radius.” However, the analogy between theorems obtained on the basis of the negation of the “fifth postulate” and formulas obtained from those of spherical geometry by replacing  $R^2$  with a negative number had been already noted by many mathematicians working on these questions (some even as early as the eighteenth century).

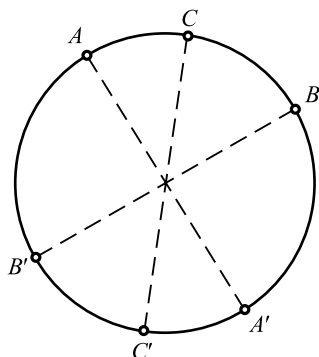
The reader should be warned that spherical geometry is entirely inconsistent with the system of axioms that we considered in Sect. 12.2. That system does not include one of the fundamental axioms of relationship: several different lines can pass through two distinct points. Indeed, infinitely many great circles pass through any two antipodal points on the sphere. In connection with this, Riemann proposed another geometry less radically different from Euclidean geometry. We shall describe it in the two-dimensional case.

For this, we shall use a description of the projective plane  $\Pi$  as the collection of all lines in three-dimensional space passing through some point  $O$ . Let us consider the sphere  $S$  with center at  $O$ . Every point  $P \in S$  together with the center  $O$  of the sphere determines a line  $l$ , that is, some point  $Q$  of the projective plane  $\Pi$ . The association  $P \rightarrow Q$  defines a mapping of the sphere  $S$  to the projective plane  $\Pi$  whereby great circles on the sphere are taken precisely to lines of  $\Pi$ . Clearly, exactly two points of the sphere are mapped to a single point  $Q \in \Pi$ : together with the point  $P$ , there is also the second point of the intersection of the line  $l$  with the sphere, that is, the antipodal point  $P'$ . But Euclidean motions taking the sphere  $S$  into itself (we might call them *motions of spherical geometry*) give certain transformations defined on the projective plane  $\Pi$  and satisfying the axioms of motion. It is possible as well to transfer the measures of lengths and angles from the sphere  $S$  to the projective plane  $\Pi$ . Then we have the analogue of Theorem 12.12 from spherical geometry.

This branch of geometry is called *elliptic geometry*.<sup>8</sup> In elliptic geometry, every pair of lines intersect, since such is the case in the projective plane. Thus there are no parallel lines. However, in “absolute geometry,” it is proved that there exists at least

---

<sup>8</sup>Elliptic geometry is sometimes called *Riemannian geometry*, but that term is usually reserved for the branch of differential geometry that studies Riemannian manifolds.

**Fig. 12.10** Elliptic geometry

one line passing through any given point  $A$  not lying on a given line  $l$  that is parallel to  $l$ . This means that in elliptic geometry, not all the axioms of “absolute geometry” are satisfied. The reason for this is easily ascertained: in elliptic geometry, there is no natural concept of “lying between.” Indeed, a great circle of the sphere  $S$  is mapped to a line  $l$  of the projective plane  $\Pi$ , where two antipodal points of the sphere ( $A$  and  $A'$ ,  $B$  and  $B'$ ,  $C$  and  $C'$ , and so on) are taken to one point of the plane  $\Pi$ . See Fig. 12.10. It is clear from the figure that in elliptic geometry, we may assume equally well that the point  $C$  does or does not lie between  $A$  and  $B$ .

Nevertheless, elliptic geometry possesses the property of “free mobility.” Moreover, one can prove (Helmholtz–Lie theorem) that among all geometries (assuming some rigorous definition of this term), only three of them—Euclidean, hyperbolic, and elliptic—possess this property.

# Chapter 13

## Groups, Rings, and Modules

### 13.1 Groups and Homomorphisms

The concept of a group is defined axiomatically, analogously to the notions of vector, inner product, and affine space. Such an abstract definition is justified by the wealth of examples of groups throughout all of mathematics.

**Definition 13.1** A *group* is a set  $G$  on which is defined an operation that assigns to each pair of elements of this set some third element; that is, there is defined a mapping  $G \times G \rightarrow G$ . The element associated with the elements  $g_1$  and  $g_2$  by this rule is called their *product* and is denoted by  $g_1 \cdot g_2$  or simply  $g_1 g_2$ . For this mapping, the following conditions must also be satisfied:

- (1) There exists an element  $e \in G$  such that for every  $g \in G$ , we have the relationships  $eg = g$  and  $ge = g$ . This element is called the *identity*.<sup>1</sup>
- (2) For each element  $g \in G$ , there exist an element  $g' \in G$  such that  $gg' = e$  and an element  $g'' \in G$  such that  $g''g = e$ . The element  $g'$  is called a *right inverse*, and the element  $g''$  is called a *left inverse* of the element  $g$ .
- (3) For every triple of elements  $g_1, g_2, g_3 \in G$ , the following relationship holds:

$$(g_1 g_2) g_3 = g_1 (g_2 g_3). \quad (13.1)$$

This last property is called *associativity*, and it is a property that we have already met repeatedly, for example in connection with the composition of mappings and matrix multiplication, and also in the construction of the exterior algebra. We considered the associative property in its most general form on p. xv, where we proved that equality (13.1) makes it possible to define the product of an *arbitrary* number of factors  $g_1 g_2 \cdots g_k$ , which then depends only on the order of the factors and not

---

<sup>1</sup>The identity element of a group is unique. Indeed, if there existed another identity element  $e' \in G$ , then by definition, we would have the equalities  $ee' = e'$  and  $ee' = e$ , from which it follows that  $e = e'$ .

on the arrangement of parentheses in the product. The reasoning given there applies, obviously, to every group.

The condition of associativity has other important consequences. From it, derives, for example, the fact that if  $g'$  is a right inverse of  $g$ , and  $g''$  is a left inverse, then

$$g''(gg') = g''e = g'', \quad g''(gg') = (g''g)g' = eg' = g',$$

from which it follows that  $g' = g''$ . Thus the left and right inverses of any given element  $g \in G$  coincide. This unique element  $g' = g''$  is called simply the *inverse* of  $g$  and is denoted by  $g^{-1}$ .

**Definition 13.2** If the number of elements belonging to a group  $G$  is finite, then the group  $G$  is called a *finite group*, and otherwise, it is called an *infinite group*. The number of distinct elements in a finite group  $G$  is called its *order* and is denoted by  $|G|$ .

Let  $M$  be an arbitrary set, and let us consider the collection of all bijective mappings between  $M$  and itself. Such mappings are also called *transformations* of the set  $M$ . In the introductory section of this book, we defined the operation of composition (that is, the sequential application) of arbitrary mappings of arbitrary sets (p. xiv). It follows from the properties proved there that the collection of all transformations of a set  $M$  together with the operation of composition forms a group, where the inverse of each transformation  $f : M \rightarrow M$  is given by the inverse mapping  $f^{-1} : M \rightarrow M$ , while the identity is obviously given by the identity mapping on the set  $M$ . Such groups are called *transformation groups*, and it is with these that the majority of applications of groups are associated.

It is sometimes necessary to consider not all the transformations of a set, but to limit our consideration to some subset. The situation that thus arises can be formulated conveniently as follows:

**Definition 13.3** A subset  $G' \subset G$  of elements of a group  $G$  is called a *subgroup* of  $G$  if the following conditions are satisfied:

- (a) For every pair of elements  $g_1, g_2 \in G'$ , their product  $g_1g_2$  is again in  $G'$ .
- (b)  $G'$  contains the identity element  $e$ .
- (c) For every  $g \in G'$ , its inverse  $g^{-1}$  is again in  $G'$ .

It is obvious that a subgroup  $G'$  is itself a group. Thus from the group of all transformations, we obtain a set of examples (indeed, the majority of examples of groups). Let us enumerate some that are met most frequently.

**Example 13.4** The following sets are groups under the operation of composition of mappings.

- 1. the set of nonsingular linear transformations of a vector space;
- 2. the set of orthogonal transformations of a Euclidean space;



3. the set of proper orthogonal transformations of a Euclidean space;
4. the set of Lorentz transformations of a pseudo-Euclidean space;
5. the set of nonsingular affine transformations of an affine space;
6. the set of projective transformations of a projective space;
7. the set of motions of an affine Euclidean space;
8. the set of motions of a hyperbolic space.

All the groups enumerated above are groups of transformations (the set  $M$  is obviously the underlying set of the given space). Let us note that in the case of vector and affine spaces, there is the crucial requirement of the nonsingularity of the linear or affine transformations that guarantees the bijectivity of each mapping and thus the existence of an inverse element for each element of the group.<sup>2</sup>

However, not all naturally occurring groups are groups of transformations. For example, with respect to the operation of addition, the set of all integers forms a group, as do the sets of the rational, real, and complex numbers, and likewise, the set of all vectors belonging to any arbitrary vector space.

Let us remark that the axioms of motion 1, 2, and 3 introduced in Sect. 12.2 can be expressed together as a single requirement, namely that the *motions form a group*.

**Example 13.5** Let us consider a finite set  $M$  consisting of  $n$  elements. A transformation  $f : M \rightarrow M$  is called a *permutation*, and the group of all permutations of the set  $M$  is called the *symmetric group of degree  $n$*  and is denoted by  $S_n$ . It is obvious that the group  $S_n$  is finite.

We considered permutations earlier, in Sect. 2.6, in connection with the notions of symmetric and antisymmetric functions, and we saw that for defining a permutation  $f : M \rightarrow M$ , one can introduce a numeration of the elements of the set  $M$ , that is, one can write the set in the form  $M = \{a_1, \dots, a_n\}$  and designate the images  $f(a_1), \dots, f(a_n)$  of all the elements  $a_1, \dots, a_n$ . Namely, let  $f(a_1) = a_{j_1}, \dots, f(a_n) = a_{j_n}$ . Then a permutation is defined by the matrix

$$A = \begin{pmatrix} 1 & 2 & \cdots & n \\ j_1 & j_2 & \cdots & j_n \end{pmatrix}, \quad (13.2)$$

where in the upper row are written in succession all the natural numbers from 1 to  $n$ , and in the lower row, under the number  $k$  stands the number  $j_k$  such that  $f(a_k) = a_{j_k}$ . Since a permutation  $f : M \rightarrow M$  is a bijective mapping, it follows that the lower row contains all the numbers from 1 to  $n$ , except that they are written in some other order. In other words,  $(j_1, \dots, j_n)$  is some permutation of the numbers  $(1, \dots, n)$ .

---

<sup>2</sup>Unfortunately, there is a certain amount of disagreement over terminology, of which the reader should be aware: above, we defined a transformation of a set as a *bijective mapping* into itself, while at the same time, a linear (or affine) transformation of a vector (or affine) space is not by definition necessarily bijective, and to have bijectivity here, it is necessary to specify that the transformations be nonsingular.

Writing a permutation in the form (13.2) allows us in particular to ascertain easily that  $|S_n| = n!$ . Let us prove this by induction on  $n$ . For  $n = 1$ , this is obvious: the group  $S_1$  contains the single permutation that is the identity mapping on the set  $M$  consisting of a single element. Let  $n > 1$ . Then by enumerating the elements of the set  $M$  in every possible way, we obtain a bijection between  $S_n$  and the set of matrices  $A$  of the form (13.2), whose first row contains the elements  $1, \dots, n$ , and the elements  $j_1, \dots, j_n$  of the second row take all possible values from 1 to  $n$ . Let  $A'$  be the matrix obtained from  $A$  by deleting its last column, containing the element  $j_n$ . Let us fix this element:  $j_n = k$ . Then the elements  $j_1, \dots, j_{n-1}$  of the matrix  $A'$  assume all possible values from the collection of the  $n - 1$  numbers  $(1, \dots, \check{k}, \dots, n)$ , where the symbol  $\check{\phantom{x}}$ , as before, denotes the omission of the corresponding element. It is clear that the set of all possible matrices  $A'$  is in bijective correspondence with  $S_{n-1}$ , and by the induction hypothesis, the number of distinct matrices  $A'$  is equal to  $|S_{n-1}| = (n - 1)!$ . But since the element  $j_n = k$  can be equal to any natural number from 1 to  $n$ , the number of distinct matrices  $A$  is equal to  $n(n - 1)! = n!$ . This gives us the equality  $|S_n| = n!$ .

Let us note that the numeration of the elements of the set  $M$  used for writing down permutations plays the same role as the introduction of coordinates (that is, a basis) in a vector space. Furthermore, the matrix (13.2) is analogous to the matrix of a linear transformation of a space, which is defined only after the choice of a basis and depends on that choice. However, for our further purposes, it will be more convenient to use concepts that are not connected with such a choice of numeration of elements.

We shall use the concept of transposition, which was introduced in Sect. 2.6 (p. 45). The definition given there can be formulated as follows. Let  $a$  and  $b$  be two distinct elements of the set  $M$ . Then a *transposition* is a permutation of the set  $M$  that interchanges the places of the elements  $a$  and  $b$  and leaves all other elements of the set  $M$  fixed. Denoting such a transposition by  $\tau_{a,b}$ , we can express this definition by the relationships

$$\tau_{a,b}(a) = b, \quad \tau_{a,b}(b) = a, \quad \tau_{a,b}(x) = x \quad (13.3)$$

for all  $x \neq a$  and  $x \neq b$ .

In this notation, Theorem 2.23 from Sect. 2.6 can be formulated as follows: *every permutation  $g$  of a finite set is the product of a finite number of transpositions*, that is,

$$g = \tau_{a_1, b_1} \tau_{a_2, b_2} \cdots \tau_{a_k, b_k}. \quad (13.4)$$

As we saw in Sect. 2.6, in relationship (13.4), the number  $k$  and the choice of elements  $a_1, b_1, \dots, a_k, b_k$  for the given permutation  $g$  are not uniquely defined. This means that for a given permutation  $g$ , the representation (13.4) is not unique. However, as was proved in Sect. 2.6 (Theorem 2.25), the parity of the number  $k$  of a permutation  $g$  is uniquely determined. Permutations for which the number  $k$  in the representation (13.4) is even are called *even*, and those for which the number  $k$  is odd are called *odd*.

**Example 13.6** The collection of all even permutations of  $n$  elements forms a subgroup of the symmetric group  $S_n$  (it obviously satisfies conditions (a), (b), (c) in the definition of a subgroup). It is called the *alternating group of degree  $n$*  and is denoted by  $A_n$ .

**Definition 13.7** Let  $g$  be an element of  $G$ . Then for every natural number  $n$ , the element  $g^n = g \cdots g$  ( $n$ -fold product) is defined. For a negative integer  $m$ , the element  $g^m$  is equal to  $(g^{-1})^{-m}$ , and for zero, we have  $g^0 = e$ .

It is easily verified that for arbitrary integers  $m$  and  $n$ , we have the relationship

$$g^m g^n = g^{m+n}.$$

From this, it is clear that the collection of elements of the form  $g^n$ , where  $n$  runs over the set of integers, forms a subgroup. It is called the *cyclic subgroup generated by the element  $g$*  and is denoted by  $\{g\}$ .

There are two cases that can occur:

- (a) All the elements  $g^n$ , as  $n$  runs through the set of integers, are distinct. In this case, we say that  $g$  is an element of *infinite order* in the group  $G$ .
- (b) For some integers  $m$  and  $n$ ,  $m \neq n$ , we have the equality  $g^m = g^n$ . Then, obviously,  $g^{m-n} = e$ . This means that there exists a natural number  $k$  (for instance  $|m - n|$ ) such that  $g^k = e$ . In this case, we say that  $g$  is an element of *finite order* in the group  $G$ .

If  $g$  is an element of finite order, then the smallest natural number  $k$  such that  $g^k = e$  is called the *order* of the element  $g$ . If for some integer  $n$ , we have  $g^n = e$ , then the number  $n$  is an integer multiple of the order  $k$  of the element  $g$ . Indeed, if such were not the case, then we could divide the number  $n$  by  $k$  with nonzero remainder:  $n = qk + r$ , where  $0 < r < k$ . From the equalities  $g^n = e$  and  $g^k = e$ , we could conclude that  $g^r = e$ , in contradiction to the definition of the order  $k$ . If in the group  $G$  there exists an element  $g$  such that  $G = \{g\}$ , then the group  $G$  is called a *cyclic group*. It is obvious that if  $G = \{g\}$  and the element  $g$  has finite order  $k$ , then  $|G| = k$ . Indeed, in this case,  $e, g, g^2, \dots, g^{k-1}$  are all the distinct elements of the group  $G$ .

Now we shall move on to discuss mappings of groups (homomorphisms), which play a role in group theory analogous to that of linear transformations of vector spaces in linear algebra. Let  $G$  and  $G'$  be any two groups, and let  $e \in G$  and  $e' \in G'$  be their identity elements.

**Definition 13.8** A mapping  $f : G \rightarrow G'$  is called a *homomorphism* if for every pair of elements  $g_1$  and  $g_2$  of the group  $G$ , we have the relationship

$$f(g_1 g_2) = f(g_1) f(g_2), \quad (13.5)$$

where it is obviously implied that on the left- and right-hand sides of equality (13.5), the juxtaposition of elements indicates the multiplication operation in the respective group (on the left, in  $G$ ; on the right, in  $G'$ ).

From equality (13.5), it is easy to derive the simplest properties of homomorphisms:

1.  $f(e) = e'$ ;
2.  $f(g^{-1}) = (f(g))^{-1}$  for every  $g \in G$ ;
3.  $f(g^n) = (f(g))^n$  for every  $g \in G$  and every integer  $n$ .

For the proof of the first property, let us set  $g_1 = g_2 = e$  in formula (13.5). Then taking into account the equality  $e = ee$ , which is obvious from the definition of the identity element, we obtain that

$$f(e) = f(ee) = f(e)f(e).$$

It remains only to multiply both sides of the relationship  $f(e) = f(e)f(e)$  by the element  $(f(e))^{-1}$  of the group  $G'$ , after which we obtain the required equality  $e' = f(e)$ . The second property follows at once from the first: setting in (13.5)  $g_1 = g$  and  $g_2 = g^{-1}$ , and taking into account the equality  $e = gg^{-1}$ , we obtain

$$e' = f(e) = f(gg^{-1}) = f(g)f(g^{-1}),$$

from which, by the definition of the inverse element, it follows that  $f(g^{-1}) = (f(g))^{-1}$ . Finally, the third property is obtained for positive  $n$  by induction from (13.5), and for negative  $n$ , it is also necessary to apply property 2.

**Definition 13.9** A mapping  $f : G \rightarrow G'$  is called an *isomorphism* if it is a homomorphism that is also a bijection. Groups  $G$  and  $G'$  are said to be *isomorphic* if there exists an isomorphism  $f : G \rightarrow G'$ . This is denoted as follows:  $G \simeq G'$ .

*Example 13.10* Assigning to each nonsingular linear transformation of a vector space  $L$  of dimension  $n$  its matrix (in some fixed basis of the space  $L$ ), we obtain an isomorphism between the group of nonsingular linear transformations of this space and the group of nonsingular square matrices of order  $n$ .

The notion of isomorphism plays the same role in group theory as the notion of isomorphism plays in the theory of vector spaces, and the notion of homomorphism plays the same role as the notion of arbitrary linear transformation (in vector spaces of arbitrary dimension). The analogy between these concepts is revealed particularly in the fact that the answer to the question whether a homomorphism  $f : G \rightarrow G'$  is an isomorphism can be formulated in terms of its *image* and *kernel*, just as was the case for linear mappings.

The *image* of a homomorphism  $f$  is the set  $f(G)$ , that is, simply the image of  $f$  as a mapping of sets  $G \rightarrow G'$ . It follows from relationship (13.5) that  $f(G)$  is a subgroup of  $G'$ . The *kernel* of a homomorphism  $f$  is the set of elements  $g \in G$  such that  $f(g) = e'$ . It is likewise not difficult to conclude from (13.5) that the kernel is a subgroup of  $G$ .

Using the notions of image and kernel, we may say that a homomorphism  $f : G \rightarrow G'$  is an isomorphism if and only if its image consists of the entire group

$G'$  and its kernel consists of only the identity element  $e \in G$ . The proof of this assertion is based on relationship (13.5) and properties 1 and 2: if for two elements  $g_1$  and  $g_2$  of a group  $G$ , we have the equality  $f(g_1) = f(g_2)$ , then through right multiplying both sides by the element  $(f(g_1))^{-1}$  of the group  $G'$ , we obtain  $e' = f(g_2)(f(g_1))^{-1} = f(g_2g_1^{-1})$ , from which it follows that  $g_2g_1^{-1} = e$ , that is,  $g_1 = g_2$ .

It is important, however, to note that the analogy between isomorphisms of groups and isomorphisms of vector spaces does not extend all that far: most of the theorems from Chap. 3 do not have suitable analogues for groups, even for finite groups. For example, one of the most important results of Chap. 3 (Theorem 3.64) states that all vector spaces of a given finite dimension are isomorphic to one another. But there exist even finite groups of a given order that are not isomorphic; see Example 13.24 on p. 484.

Another property of groups is related to whether the product of elements in a group depends on the order in which they are multiplied. In the definition of a group, no condition of this sort was imposed, and therefore, we may assume that in general,  $g_1g_2 \neq g_2g_1$ . Very frequently, such is the case. For example, nonsingular square matrices of a given order  $n$  with the standard operation of matrix multiplication form a group, and as the example presented in Sect. 2.9 on p. 64 shows, already for  $n = 2$ , it is generally the case that  $AB \neq BA$ .

**Definition 13.11** If in a group  $G$  the equality  $g_1g_2 = g_2g_1$  holds for every pair of elements  $g_1, g_2 \in G$ , then  $G$  is called a *commutative group* or, more usually, an *abelian group*.<sup>3</sup>

For example, the groups of integers, rational numbers, real numbers, and complex numbers with the operation of addition are all abelian. Likewise, a vector space is an abelian group with respect to the operation of vector addition. It is easy to see that every cyclic group is abelian.

Let us present one result that holds for all finite groups but that is especially easy to prove (and we shall use it frequently in the sequel) for abelian groups.

**Lemma 13.12** *For every finite abelian group  $G$ , the order of each of its elements divides the order of the group.*

*Proof* Let us denote by  $g_1, g_2, \dots, g_n$  the complete set of elements of  $G$  (so we obviously have  $n = |G|$ ), and let us right multiply each of them by some element  $g \in G$ . The elements thus obtained,  $g_1g, g_2g, \dots, g_ng$ , will again all be distinct. Indeed, given the equality  $g_i g = g_j g$ , right multiplying both sides by  $g^{-1}$  yields the equality  $g_i = g_j$ . Since the group  $G$  contains  $n$  elements altogether, it follows that the elements  $g_1g, g_2g, \dots, g_ng$  are the same as the elements  $g_1, g_2, \dots, g_n$ , though perhaps arranged in some other order:

$$g_1g = g_{i_1}, \quad g_2g = g_{i_2}, \quad \dots, \quad g_ng = g_{i_n}.$$

---

<sup>3</sup>Named in honor of the Norwegian mathematician Niels Henrik Abel (1802–1829).

On multiplying these equalities, we obtain

$$(g_1g)(g_2g) \cdots (g_ng) = g_{i_1}g_{i_2} \cdots g_{i_n}. \quad (13.6)$$

Since the group  $G$  is abelian, we have

$$(g_1g)(g_2g) \cdots (g_ng) = g_1g_2 \cdots g_ng^n,$$

and since  $g_{i_1}, g_{i_2}, \dots, g_{i_n}$  are the same elements  $g_1, g_2, \dots, g_n$ , then setting  $h = g_1g_2 \cdots g_n$ , we obtain from (13.6) the equality  $hg^n = h$ . Left multiplying both sides of the last equality by  $h^{-1}$ , we obtain  $g^n = e$ . As we saw above, it then follows that the order of the element  $g$  divides the number  $n = |G|$ .  $\square$

**Definition 13.13** Let  $H_1, H_2, \dots, H_r$  be subgroups of  $G$ . The group  $G$  is called the *direct product* of the subgroups  $H_1, H_2, \dots, H_r$  if for all elements  $h_i \in H_i$  and  $h_j \in H_j$  from distinct subgroups, we have the relationship  $h_ih_j = h_jh_i$ , and every element  $g \in G$  can be represented in the form

$$g = h_1h_2 \cdots h_r, \quad h_i \in H_i, i = 1, 2, \dots, r,$$

and for each element  $g \in G$ , such a representation is unique. The fact that the group  $G$  is a direct product of subgroups  $H_1, H_2, \dots, H_r$  is denoted by

$$G = H_1 \times H_2 \times \cdots \times H_r. \quad (13.7)$$

In the case of abelian groups, a different terminology is usually used, related to the majority of examples of interest. Namely, the operation defined on the group is called *addition* instead of multiplication, and it is denoted not by  $g_1g_2$ , but by  $g_1 + g_2$ . In keeping with this notation, the identity element is called the *zero element* and is denoted by 0, and not by  $e$ . The inverse element is called the *negative* or *additive inverse* and is denoted not by  $g^{-1}$ , but by  $-g$ , and the exponential notation  $g^n$  is replaced by the *multiplicative* notation  $ng$ , which is defined similarly:  $ng = g + \cdots + g$  ( $n$ -fold sum) if  $n > 0$ , by  $ng = (-g) + \cdots + (-g)$  ( $n$ -fold sum) if  $n < 0$ , and by  $ng = 0$  if  $n = 0$ . The definition of homomorphism remains exactly the same in this case, where it is required only to replace in formula (13.5) the symbol for the group operation:

$$f(g_1 + g_2) = f(g_1) + f(g_2).$$

Properties 1–3 here take the following form:

1.  $f(0) = 0'$ ;
2.  $f(-g) = -f(g)$  for all  $g \in G$ ;
3.  $f(ng) = nf(g)$  for all  $g \in G$  and for every integer  $n$ .

This terminology agrees with the example of the set of integers and, in the terminology we employed earlier, the example of vectors that form an abelian group with respect to the operation of addition.

In the case of abelian groups (with the operation of addition), instead of the direct product of subgroups  $H_1, H_2, \dots, H_r$  one speaks of their *direct sum*. Then the definition of the direct sum reduces to the condition that every element  $g \in G$  can be represented in the form

$$g = h_1 + h_2 + \dots + h_r, \quad h_i \in H_i, i = 1, 2, \dots, r,$$

and that for each element  $g \in G$ , the representation is unique. It is obvious that this last requirement is equivalent to the requirement that the equality  $h_1 + h_2 + \dots + h_r = 0$  be possible only if  $h_1 = 0, h_2 = 0, \dots, h_r = 0$ . That a group  $G$  is the direct sum of subgroups  $H_1, H_2, \dots, H_r$  is denoted by

$$G = H_1 \oplus H_2 \oplus \dots \oplus H_r. \quad (13.8)$$

It is obvious that in both cases (13.7) and (13.8), the order of the group  $G$  is equal to

$$|G| = |H_1| \cdot |H_2| \cdot \dots \cdot |H_r|.$$

In perfect analogy to how things were done in Sect. 3.1 for vector spaces, we may define the direct product (or direct sum) of groups that in general are not originally the subgroups of any particular group and that even, perhaps, are of completely different natures from one another.

*Example 13.14* If we map every orthogonal transformation  $\mathcal{U}$  of a Euclidean space to its determinant  $|\mathcal{U}|$ , which, as we know, is equal to  $+1$  or  $-1$ , we obtain a homomorphism of the group of orthogonal transformations into the symmetric group  $S_2$  of order 2. If we map every Lorentz transformation  $\mathcal{U}$  of a pseudo-Euclidean space to the pair of numbers  $\varepsilon(\mathcal{U}) = (|\mathcal{U}|, \nu(\mathcal{U}))$ , defined in Sect. 7.8, we obtain a homomorphism of the group of Lorentz transformations into the group  $S_2 \times S_2$ .

*Example 13.15* Let  $(V, L)$  be an affine Euclidean space of dimension  $n$  and  $G$  the group of its motions. Then the assertion of Theorem 8.37 can be formulated as the equality  $G = T_n \times O_n$ , where  $T_n$  is the group of translations of the space  $V$ , and  $O_n$  is the group of orthogonal transformations of the space  $L$ . Let us note that  $T_n \simeq L$ , where  $L$  is understood as a group under the operation of vector addition. Indeed, let us define the mapping  $f : T_n \rightarrow L$  that to each translation  $\mathcal{T}_a$  by the vector  $a$  assigns this vector  $a$ . Obviously, the mapping  $f$  is bijective, and by virtue of the property  $\mathcal{T}_a \mathcal{T}_b = \mathcal{T}_{a+b}$ , it is an isomorphism. Thus Theorem 8.37 can be formulated as the relationship  $G \simeq L \times O_n$ .

## 13.2 Decomposition of Finite Abelian Groups

Later in this chapter we shall restrict our attention to the study of finite groups. The highest goal in this area of group theory is to find a construction that gives a

description of all finite groups. But such a goal is far from accessible; at least at present, we are far from attaining it. However, for finite *abelian* groups, the answer to this question turns out to be unexpectedly simple. Moreover, both the answer and its proof are very similar to Theorem 5.12 on the decomposition of a vector space as a direct sum of cyclic subspaces. For the proof, we shall require the following lemmas.

**Lemma 13.16** *Let  $B$  be a subgroup of  $A$ , and  $a$  an element of the group  $A$  of order  $k$ . If there exists a number  $m \in \mathbb{N}$  relatively prime to  $k$  such that  $ma \in B$ , then  $a$  is an element of  $B$ .*

*Proof* Since the numbers  $m$  and  $k$  are relatively prime, there exist integers  $r$  and  $s$  such that  $kr + ms = 1$ . Multiplying  $ma$  by  $s$  and adding  $kra$  to the result (which is equal to zero, since  $k$  is the order of the element  $a$ ), we obtain  $a$ . But  $sma = s(ma)$  belongs to the subgroup  $B$ . From this, it follows that  $a$  is also an element of  $B$ .  $\square$

**Lemma 13.17** *If  $A = \{a\}$  is a cyclic group of order  $n$ , and we set  $b = ma$ , where  $m \in \mathbb{N}$  is relatively prime to  $n$ , then the cyclic subgroup  $B = \{b\}$  generated by the element  $b$  coincides with  $A$ .*

*Proof* Since  $a \in A$ , we have by Lemma 13.12 that the order  $k$  of the element  $a$  divides the order of the group  $A$ , which is equal to  $n$ , and the relative primality of the numbers  $m$  and  $n$  implies the relative primality of the numbers  $k$  and  $m$ . From Lemma 13.16, it follows that  $a \in B$ , which means that  $A \subset B$ , and since we obviously have also  $B \subset A$ , we obtain the required equality  $B = A$ .  $\square$

**Corollary 13.18** *Under the assumptions of Lemma 13.17, every element  $c \in A$  can be expressed in the form*

$$c = md, \quad d \in A, m \in \mathbb{Z}. \quad (13.9)$$

Indeed, if in the notation of Lemma 13.17, the group  $A$  is the group  $\{b\}$ , then the element  $c$  has the form  $kb$ , and since  $b = ma$ , we obtain equality (13.9) in which  $d = ka$ .

**Definition 13.19** A subgroup  $B$  of a group  $A$  is said to be *maximal* if  $B \neq A$  and  $B$  is contained in no subgroup other than  $A$ .

It is obvious that there exist maximal subgroups in every finite group that consists of more than just a single element. Indeed, beginning with the identity subgroup (that is, the subgroup consisting of a single element), we can include it, if it is not itself maximal, in some subgroup  $B_1$  different from  $A$ . If in  $B_1$  we have not yet obtained a maximal subgroup, then we can include it in some subgroup  $B_2$  different from  $A$ . Continuing this process, we eventually can go no further, since all the subgroups  $B_1, B_2, \dots$  are contained in the finite group  $A$ . The last subgroup



obtained when we stop the process will be maximal. We remark that we do not assert (nor is it true) that the maximal subgroup we have constructed is unique.

**Lemma 13.20** *For every maximal subgroup  $B$  of a finite abelian group  $A$ , there exists an element  $a \in A$  not belonging to  $B$  such that the smallest number  $m \in \mathbb{N}$  for which  $ma$  belongs to  $B$  is prime, and every element  $x \in A$  can be represented in the form*

$$x = ka + b, \quad (13.10)$$

for  $k$  an integer,  $b \in B$ .

Later, we shall denote the prime number  $m$  that appears in Lemma 13.20 by  $p$ .

*Proof of Lemma 13.20* Let us take as  $a$  any element of the group  $A$  not belonging to the subgroup  $B$ . The collection of all elements of the form  $ka + b$ , where  $k$  is an arbitrary integer and  $b$  an arbitrary element of  $B$ , obviously forms a subgroup containing  $B$  (it is easy to see that  $B$  consists of elements  $x$  such that in the representation  $x = ka + b$ , the number  $k$  is equal to 0). It is obvious that this subgroup does not coincide with  $B$ , since it contains the element  $a$  (for  $k = 1$  and  $b = 0$ ), and this means, in view of the maximality of the subgroup  $B$ , that it coincides with  $A$ . From this follows the representation (13.10) for every element  $x$  in the group  $A$ .

It remains to prove that for some prime number  $p$ , the element  $pa$  belongs to  $B$ . Since the element  $a$  is of finite order, we must have  $na = 0$  for some  $n > 0$ . In particular,  $na \in B$ . Let us take the smallest  $m \in \mathbb{N}$  for which  $ma \in B$  and prove that it is prime.

Suppose that such is not the case, and that  $p$  is a prime divisor of  $m$ . Then  $m = pm_1$  for some integer  $m_1 < m$ . Let us set  $a_1 = m_1a$ . As we have seen, the collection of all elements of the form  $ka_1 + b$  (for arbitrary integer  $k$  and  $b \in B$ ) forms a subgroup of the group  $A$  containing  $B$ . If the element  $a_1$  were contained in  $B$ , then that would contradict the choice of  $m$  as the *smallest* natural number such that  $ma \in B$ . This means that  $a_1 \notin B$ , and in view of the maximality of the subgroup  $B$ , the subgroup that we constructed of elements of the form  $ka_1 + b$  coincides with  $A$ . In particular, it contains the element  $a$ , that is,  $a = ka_1 + b$  for some  $k$  and  $b$ . From this, it follows that  $pa = kpa_1 + pb$ . But  $pa_1 = pm_1a = ma \in B$ , and since  $pb \in B$ , this means that  $pa \in B$ , which contradicts the minimality of  $m$ . This means that the assumption that  $m$  has prime divisors less than  $m$  is false, and so  $m = p$  is a prime number.  $\square$

*Remark 13.21* We chose as  $a$  an arbitrary element of the group  $A$  not contained in  $B$ . In particular, in place of  $a$ , we could as well choose any element  $a' = a + b$ , where  $b \in B$ . Indeed, from  $a = a' - b$  and  $a' \in B$  it would follow that we would also have  $a \in B$ .

We can now state the fundamental theorem of abelian groups.

**Theorem 13.22** *Every finite abelian group is the direct sum of cyclic subgroups whose orders are equal to powers of prime numbers.*

Thus, the theorem asserts that every finite abelian group  $A$  has the decomposition

$$A = A_1 \oplus \cdots \oplus A_r, \quad (13.11)$$

where the subgroups  $A_i$  are cyclic, that is,  $A_i = \{a_i\}$ , and their orders are powers of prime numbers, that is,  $|A_i| = p_i^{m_i}$ , where  $p_i$  are prime numbers.

*Proof of Theorem 13.22* Our proof is by induction on the order of the group  $A$ . For the group of order 1, the theorem is obvious. Therefore, to prove the theorem for a group  $A$ , we may assume that it has been proved for all subgroups  $B \subset A$ ,  $B \neq A$ , since for an arbitrary subset  $B \subset A$  with  $B \neq A$ , the number of elements of  $B$  is less than  $|A|$ .

In particular, let  $B$  be a maximal subgroup of the group  $A$ . By the induction hypothesis, the theorem is valid for this subgroup, and it therefore has the decomposition

$$B = C_1 \oplus \cdots \oplus C_r, \quad (13.12)$$

in which the  $C_i$  are cyclic subgroups each of which has order the power of a prime number:

$$C_i = \{c_i\}, \quad p_i^{m_i} c_i = 0.$$

Lemma 13.20 holds for the subgroup  $B$ ; let  $a \in A$ ,  $a \notin B$ , be the element provided for in the formulation of this lemma. By hypothesis, every element  $x \in B$  can be represented in the form

$$x = k_1 c_1 + \cdots + k_r c_r.$$

In particular, this holds for the element  $b = pa$  (in the notation of Lemma 13.20):

$$pa = k_1 c_1 + \cdots + k_r c_r.$$

Let us select the terms  $k_i c_i$  in this decomposition that can be written in the form  $pd_i$ , where  $d_i \in C_i$ . These are first of all, the terms  $k_i c_i$  for  $i$  such that  $p_i \neq p$ . This follows from Corollary 13.18. Moreover, all elements of the form  $k_i c_i$  possess this property if  $p_i = p$  and  $k_i$  is divisible by  $p$ . Let the chosen elements be  $k_i c_i$ ,  $i = 1, \dots, s-1$ . Then for the remaining elements  $k_i c_i$ ,  $i = s, \dots, r$ , we have  $p_i = p$  and  $k_i$  is not divisible by  $p$ . Setting

$$k_i c_i = pd_i, \quad d_i \in C_i, i = 1, \dots, s-1, \quad d_1 + \cdots + d_{s-1} = d, \quad (13.13)$$

we obtain

$$pa = pd + k_s c_s + \cdots + k_r c_r.$$

We can now use the freedom in the choice of the element  $a \in A$ , which was mentioned in Remark 13.21, and take instead of  $a$ , the element  $a' = a - d$ , since  $d \in B$  in view of formula (13.13). We then have

$$pa' = k_s c_s + \cdots + k_r c_r. \quad (13.14)$$

There are now two possible cases.

*Case 1.* The number  $s - 1$  is equal to  $r$ , and then equality (13.14) gives

$$pa' = 0.$$

In this case, the group  $A$  decomposes as a direct sum of cyclic subgroups as follows:

$$A = C_1 \oplus \cdots \oplus C_r \oplus C_{r+1},$$

where  $C_{r+1} = \{a'\}$  is a subgroup of order  $p$ .

Indeed, Lemma 13.20 asserts that every element  $x \in A$  can be represented in the form  $ka' + b$ , and since in view of (13.12), the element  $b$  can be represented in the form

$$b = k_1 c_1 + \cdots + k_r c_r,$$

it follows that  $x$  has the form

$$x = k_1 c_1 + \cdots + k_r c_r + ka'. \quad (13.15)$$

This proves the first condition in the definition of a direct sum.

Let us prove the uniqueness of representation (13.15). For this, it suffices to prove that the equality

$$k_1 c_1 + \cdots + k_r c_r + ka' = 0 \quad (13.16)$$

is possible only for  $k_1 c_1 = \cdots = k_r c_r = ka' = 0$ . Let us rewrite (13.16) in the form

$$ka' = -k_1 c_1 - \cdots - k_r c_r. \quad (13.17)$$

This means that the element  $ka'$  belongs to  $B$ . If the number  $k$  were not divisible by  $p$ , then  $k$  and  $p$  would be relatively prime, since the element  $a'$  has order  $p$ , and by Lemma 13.16, we would then obtain that  $a' \in B$ . But this contradicts the choice of the element  $a$  and the construction of the element  $a'$ . This means that  $p$  must divide  $k$ , and since  $pa' = 0$ , it follows that we also have  $ka' = 0$ . Thus equality (13.17) is reduced to  $k_1 c_1 + \cdots + k_r c_r = 0$ , and from the fact that the group  $B$  is the direct sum of subgroups  $C_1, \dots, C_r$ , we obtain that  $k_1 c_1 = 0, \dots, k_r c_r = 0$ .

*Case 2.* The number  $s - 1$  is less than  $r$ . Let us set  $k_s c_s = d_s, \dots, k_r c_r = d_r$ , and for  $i = 1, \dots, s - 1$ , let us set  $c_i = d_i$ . By Lemma 13.17, the element  $d_i$  generates the same cyclic subgroup  $C_i$  as  $c_i$ . For  $i \leq s - 1$ , this assertion is a tautology, and for  $i > s - 1$ , it follows from the fact that the numbers  $k_i$  are by assumption not

divisible by  $p$ , and  $p^{m_i}c_i = 0$  for all  $i \geq s$ . Equality (13.14) can then be rewritten as follows:

$$pa' = d_s + \cdots + d_r. \quad (13.18)$$

Let  $m_s \leq \cdots \leq m_r$ . Let us denote by  $C'_r$  the cyclic group generated by the element  $a'$ , that is, let us set  $C'_r = \{a'\}$ . Let us prove that the order of the element  $a'$ , and therefore the order of the group  $C'_r$ , is equal to  $p^{m_r+1}$ :

$$|C'_r| = p^{m_r+1}. \quad (13.19)$$

Indeed, in view of (13.18), we have

$$p^{m_r+1}a' = p^{m_r}d_s + \cdots + p^{m_r}d_r = 0,$$

since  $p^{m_i}d_i = 0$ ,  $m_i \leq m_r$ . On the other hand, in view of relationship (13.18), we have

$$p^{m_r}a' = p^{m_r-1}d_s + \cdots + p^{m_r-1}d_r \neq 0,$$

since  $p^{m_r-1}d_r \neq 0$ , and in view of (13.12), the sum of the elements  $p^{m_r-1}d_i \in C_i$  cannot equal 0 if at least one term is not equal to 0. This proves (13.19).

Now let us prove that

$$A = C_1 \oplus \cdots \oplus C_{r-1} \oplus C'_r, \quad (13.20)$$

that is, that every element  $x \in A$  can be uniquely represented in the form

$$x = y_1 + \cdots + y_{r-1} + y'_r, \quad y_1 \in C_1, \dots, y_{r-1} \in C_{r-1}, y'_r \in C'_r. \quad (13.21)$$

First let us prove the possibility of representation (13.21). Since every element  $x \in A$  can be represented in the form  $ka' + b$ ,  $b \in B$ , it suffices to prove that it is possible to represent separately  $a'$  and an arbitrary element  $b \in B$  in the form (13.21). This is obvious for an element  $a'$ , since it belongs to the cyclic group  $C'_r = \{a'\}$ . As for elements of  $B$ , each  $b \in B$  can be represented in the form

$$b = k_1d_1 + \cdots + k_rd_r,$$

according to formula (13.12) and in view of the fact that  $C_i = \{d_i\}$ . Therefore, it suffices to prove that each of the elements  $d_i$  can be represented in the form (13.21). For  $d_1, \dots, d_{r-1}$ , this is obvious, since

$$d_i \in C_i = \{d_i\}, \quad i = 1, \dots, r-1.$$

Finally, in view of (13.18), we have

$$d_r = -d_s - \cdots - d_{r-1} + pa',$$

and this is the representation of the element  $d_r$  that we need.

Let us now prove the uniqueness of representation (13.21). For this, it suffices to prove that the equality

$$k_1 d_1 + \cdots + k_{r-1} d_{r-1} + k_r a' = 0 \quad (13.22)$$

is possible only for  $k_1 d_1 = \cdots = k_r a' = 0$ . Let us suppose that  $k_r$  is relatively prime to  $p$ . Then

$$k_r a' = -k_1 d_1 - \cdots - k_{r-1} d_{r-1},$$

and in view of the fact that  $p^{m_r+1} a' = 0$ , we obtain by Lemma 13.16 that  $a' \in B$ . But the element  $a \in A$  was chosen as an element not belonging to the subgroup  $B$ . This means that the element  $a'$  also does not belong to  $B$ .

Let us now consider the case in which the number  $k_r$  is divisible by  $p$ . Let  $k_r = pl$ . Then

$$pla' = -k_1 d_1 - \cdots - k_{r-1} d_{r-1}.$$

Let us replace  $pa'$  on the left-hand side of this relationship by the expression  $d_s + \cdots + d_r$  on the basis of equality (13.18). On transferring all terms to the left-hand side, we obtain

$$ld_s + \cdots + ld_r + k_1 d_1 + \cdots + k_{r-1} d_{r-1} = 0.$$

From the fact that by hypothesis, the group  $B$  is the direct sum of groups  $C_1, \dots, C_r$ , it follows that in this equality,  $ld_r = 0$ . Since the order of the element  $d_r$  is equal to  $p^{m_r}$ , this is possible only if  $p^{m_r}$  divides  $l$ , and this means that  $p^{m_r+1}$  divides  $k_r$ . But we have seen that the order of the element  $a'$  is equal to  $p^{m_r+1}$ , and this means that  $k_r a' = 0$ . Then it follows from equality (13.22) that  $k_1 d_1 + \cdots + k_{r-1} d_{r-1} = 0$ . And since by the induction hypothesis, the group  $B$  is the direct sum of the groups  $C_1, \dots, C_r$ , it follows that  $k_1 d_1 = \cdots = k_{r-1} d_{r-1} = 0$ . This completes the proof of the theorem.  $\square$

### 13.3 The Uniqueness of the Decomposition

The theorem on the uniqueness of the Jordan normal form has an analogue in the theory of finite abelian groups.

**Theorem 13.23** *For different decompositions of the finite abelian group  $A$  into a direct sum of cyclic subgroups whose orders are prime powers, whose existence is established in Theorem 13.22,*

$$A = A_1 \oplus \cdots \oplus A_r, \quad |A_i| = p_i^{m_i}, \quad (13.23)$$

*the orders  $p_i^{m_i}$  of the cyclic subgroups  $A_i$  are unique. In other words, if*

$$A = A'_1 \oplus \cdots \oplus A'_s$$

is another such decomposition, then  $s = r$ , and the subgroups  $A'_i$  can be reordered in such a way that the equality  $|A'_i| = |A_i|$  is satisfied for all  $i = 1, \dots, r$ .

*Proof* We shall show how the orders of the cyclic subgroups in the decomposition (13.23) are uniquely determined by the group  $A$  itself. For any natural number  $k$ , let us denote by  $kA$  the collection of elements  $a$  of the group  $A$  that can be represented in the form  $a = kb$ , where  $b$  is some element of this group. It is obvious that the collection of elements  $kA$  forms a subgroup of the group  $A$ . Let us prove that the orders  $|kA|$  of these subgroups (for various  $k$ ) determine the orders of the cyclic groups  $|A_i|$  in the decomposition (13.23).

Let us consider an arbitrary prime number  $p$  and analyze the case that  $k$  is a power of a prime number  $p$ , that is,  $k = p^i$ . Let us factor the order  $|p^i A|$  of the group  $p^i A$  into a product of a power of  $p$  and numbers  $n_i$  relatively prime to  $p$ :

$$|p^i A| = p^{r_i} n_i, \quad (n_i, p) = 1. \quad (13.24)$$

On the other hand, for a prime number  $p$ , let us denote by  $l_i$  the number of subgroups  $A_i$  of order  $p^i$  appearing in the decomposition (13.23). We shall present an explicit formula that expresses the numbers  $l_i$  in terms of  $r_i$ . Since these latter numbers are determined only by the group  $A$ , it follows that the numbers  $l_i$  also do not depend on the decomposition (13.23) (in particular, they are equal to zero if and only if all prime numbers  $p_i$  for which  $|A_i| = p_i^{m_i}$  differ from  $p$ ).

First of all, let us calculate the order of the group  $A$  in another way. Let us note that  $A = p^0 A$ , so that this is the case  $i = 0$ . The definition of the number  $l_i$  shows that in the decomposition (13.23), we have  $l_1$  groups of order  $p$ ,  $l_2$  groups of order  $p^2$ ,  $\dots$ , and the remaining groups have orders relatively prime to  $p$ . Hence it follows that

$$|A| = p^{l_1} p^{2l_2} \dots n_0, \quad (n_0, p) = 1.$$

Let us set

$$|A| = p^{r_0} n_0, \quad (n_0, p) = 1.$$

Then we can write the relationship above in the form

$$l_1 + 2l_2 + 3l_3 + \dots = r_0. \quad (13.25)$$

Now let us consider the case that  $k = p^i > 1$ , that is, the number  $i$  is greater than 0. First of all, it is obvious that for every natural number  $k$ , it follows from (13.23) that

$$kA = kA_1 \oplus kA_2 \oplus \dots \oplus kA_r.$$

It is obvious that all properties of a direct sum are satisfied.

Now, as in the case examined above, let us calculate the order of the group  $p^i A$  in another way. It is obvious that  $|p^i A| = |p^i A_1| \dots |p^i A_r|$ . If for some  $j$ , we have  $|A_j| = p_j^{m_j}$  and  $p_j \neq p$ , then Lemma 13.17 shows that  $p^i A_j = A_j$ , and we have

$|p^i A_j| = |A_j| = p_j^{m_j}$ , which is relatively prime to  $p$ . Thus in the decomposition  $|p^i A| = |p^i A_1| \cdots |p^i A_r|$ , all the factors  $|p^i A_j|$ , where  $|A_j| = p_j^{m_j}$  and  $p_j \neq p$ , together give a number that is relatively prime to  $p$ , and in formula (13.24), they make no contribution to the number  $r_i$ . It remains to consider the case  $p_j = p$ . Since  $A_j$  is a cyclic group, it follows that  $A_j = \{a_j\}$ . It is then clear that  $p^i A_j = \{p^i a_j\}$ . Let us find the order of the element  $p^i a_j$ . Since  $p^{m_j} a_j = 0$ , we have  $p^{m_j-i}(p^i a_j) = 0$  if  $i \leq m_j$ , and  $p^i a_j = 0$  if  $i = m_j$ .

Let us prove that  $p^{m_j-i}$  is precisely the same as the order of the element  $p^i a_j$ . Let this order be equal to some number  $s$ . Then  $s$  must divide  $p^{m_j-i}$ , which means that it is of the form  $p^t$ . If  $t < m_j - i$ , then the equality  $p^t(p^i a_j) = 0$  would show that  $p^{t+i} a_j = 0$ , that is, that the element  $a_j$  had order less than  $p^{m_j}$ . This means that  $|p^i A_j| = p^{m_j-i}$  for  $i \leq m_j$ . The fact that  $p^i A_j = 0$  for  $i \geq m_j$  (which means that  $|p^i A_j| = 1$ ) is obvious.

We can now literally repeat the argument that we used earlier. We see that in the decomposition

$$p^i A = p^i A_1 \oplus p^i A_2 \oplus \cdots \oplus p^i A_r,$$

subgroups of order  $p$  occur when  $m_j - i = 1$ , that is,  $m_j = i + 1$ , and this means that in our adopted notation, they occur  $l_{i+1}$  times. Likewise, the subgroups of order  $p^2$  occur when  $m_j = i + 2$ , that is,  $l_{i+2}$  times, and so on. Moreover, certain subgroups will have order relatively prime to  $p$ . This means that

$$|p^i A| = p^{l_{i+1}} p^{2l_{i+2}} \cdots n_i, \quad \text{where } (n_i, p) = 1.$$

In other words, in accordance with our previous notation, we have

$$l_{i+1} + 2l_{i+2} + \cdots = r_i. \quad (13.26)$$

In particular, formula (13.25) is obtained from (13.26) for  $i = 0$ .

If we now subtract from each formula (13.26) the following one, we obtain that for all  $i = 1, 2, \dots$ , we have the equalities

$$l_i + l_{i+1} + \cdots = r_{i-1} - r_i.$$

Repeating the same process, we obtain

$$l_i = r_{i-1} - 2r_i + r_{i+1}.$$

These relationships prove Theorem 13.23. □

Theorems 13.22 and 13.23 make it easy to give the number of distinct (up to isomorphism) finite abelian groups of a given order.

*Example 13.24* Suppose, for example, that we would like to determine the number of distinct abelian groups of order  $p^3 q^2$ , where  $p$  and  $q$  are distinct prime numbers. Theorem 13.22 shows that such a group can be represented in the form

$$A = C_1 \oplus \cdots \oplus C_s,$$

where  $C_i$  are cyclic groups whose orders are prime powers. From this decomposition, it follows that

$$|A| = |C_1| \cdots |C_s|.$$

In other words, among the groups  $C_i$ , there is either one cyclic group of order  $p^3$ , or one of order  $p^2$  and one of order  $p$ , or three of order  $p$ . And likewise, there is one of order  $q^2$  or two of order  $q$ . Combining all these possibilities (three for groups of order  $p^i$  and two for groups of order  $q^j$ ), we obtain six variants. Theorem 13.23 guarantees that of the six groups thus obtained, none is isomorphic to any of the others.

### 13.4 Finitely Generated Torsion Modules over a Euclidean Ring\*

The proofs of the theorem on finite abelian groups and the theorem on Jordan normal form (just like the proofs of the corresponding uniqueness theorems) are so obviously parallel to each other that they surely are special cases of some more general theorems. This is indeed the case, and the main goal of this chapter is the proof of these general theorems. For this, we shall need two abstract (that is, defined axiomatically) notions.

**Definition 13.25** A *ring* is a set  $R$  on which are defined two operations (that is, two mappings  $R \times R \rightarrow R$ ), one of which is called *addition* (for which an element that is the image of two elements  $a \in R$  and  $b \in R$  is called their *sum* and is denoted by  $a + b$ ), and the second of which is *multiplication* (the element that is the image of  $a \in R$  and  $b \in R$  is called their *product* and is denoted by  $ab$ ). For these operations of addition and multiplication, the following conditions must be satisfied:

- (1) With respect to the operation of addition, the ring is an abelian group (the identity element is denoted by 0).
- (2) For all  $a, b, c \in R$ , we have

$$a(b + c) = ab + ac, \quad (b + c)a = ba + ca.$$

- (3) For all  $a, b, c \in R$ , the associative property holds:

$$a(bc) = (ab)c.$$

In the sequel, we shall denote a ring by the letter  $R$  and assume that it has a multiplicative identity, that is, that it contains an element, which we shall denote by 1, satisfying the condition

$$a \cdot 1 = 1 \cdot a = a \quad \text{for all } a \in R.$$

In this chapter, we shall be considering only *commutative* rings, that is, it will be assumed that

$$ab = ba \quad \text{for all } a, b \in R.$$



We have already encountered the most important special case of a ring, namely an algebra, in connection with the construction of the exterior algebra of a vector space, in Chap. 10. Let us recall that an algebra is a ring that is a vector space, where, of course, consistency of the notions entering into these definitions is assumed. This means that for every scalar  $\alpha$  (in the field over which the vector space in question is defined) and for all elements  $a, b$  of the ring  $R$ , we have the equality  $(\alpha a)b = \alpha(ab)$ . On the other hand, we are quite familiar with an example of a ring that is not an algebra in any natural sense, namely the ring of integers  $\mathbb{Z}$  with the usual arithmetic operations of addition and multiplication.

Let us note a connection among the concepts we have introduced. If all nonzero elements of a commutative ring form a group with respect to the operation of multiplication, then such a ring is called a *field*. We assume that the reader is familiar with the simplest properties of fields and rings.

The concept that generalizes both the concept of vector space (over some field  $\mathbb{K}$ ) with a linear transformation given on it and that of an abelian group is that of a *module*.

**Definition 13.26** An abelian group  $M$  (its operation is written as addition) is a *module*  $M$  over a ring  $R$  if there is defined an additional operation of multiplication of the elements of the ring  $R$  by elements of the module  $M$  that produces elements of the module that have the following properties:

$$\begin{aligned} a(\mathbf{m} + \mathbf{n}) &= a\mathbf{m} + a\mathbf{n}, \\ (a + b)\mathbf{m} &= a\mathbf{m} + b\mathbf{m}, \\ (ab)\mathbf{m} &= a(b\mathbf{m}), \\ 1\mathbf{m} &= \mathbf{m}, \end{aligned}$$

for all elements  $a, b \in R$  and all elements  $\mathbf{m}, \mathbf{n} \in M$ .

For convenience, we shall denote the elements of the ring using ordinary letters  $a, b, \dots$ , and elements of the module using boldface letters:  $\mathbf{m}, \mathbf{n}, \dots$ .

*Example 13.27* An example of a module that we have encountered repeatedly is that of a vector space over an arbitrary field  $\mathbb{K}$  (here the ring  $R$  is the field  $\mathbb{K}$ ). On the other hand, every abelian group  $G$  is a module over the ring of integers  $\mathbb{Z}$ : the operation defined on it of integral multiplication  $k\mathbf{g}$  for  $k \in \mathbb{Z}$  and  $\mathbf{g} \in G$  obviously possesses all the required properties.

*Example 13.28* Let  $L$  be a vector space (real, complex, or over an arbitrary field  $\mathbb{K}$ ) and let  $\mathcal{A} : L \rightarrow L$  be a fixed linear transformation. Then we may consider  $L$  as a module over the ring  $R$  of polynomials in the single variable  $x$  (real, complex, or over a field  $\mathbb{K}$ ), assuming, as we did earlier, for a polynomial  $f(x) \in R$  and vector  $\mathbf{e} \in L$ ,

$$f(x)\mathbf{e} = f(\mathcal{A})(\mathbf{e}). \quad (13.27)$$

It is easily verified that all the properties appearing in the definition of a module are satisfied.

Our immediate objective will be to find a restriction of the general notion of module that covers vector spaces and abelian groups and then to prove theorems for these that generalize Theorems 5.12 and 13.22.

These two examples—the ring of integers  $\mathbb{Z}$  and the ring of polynomials in a single complex variable (for simplicity, we shall restrict our attention to the special case  $\mathbb{K} = \mathbb{C}$ , but many results are valid in the general case)—have many similar properties, the most important of which is the uniqueness of the decomposition into irreducible factors, that is, prime numbers in the case of the ring of integers, and linear polynomials in the case of the ring of polynomials with complex coefficients. Both of these properties, in turn, derive from a single property: the possibility of division with remainder, which we shall introduce in the definition of certain rings for which it is possible to generalize the reasoning from previous sections.

**Definition 13.29** A ring  $R$  is called a *Euclidean ring* if

$$ab \neq 0 \quad \text{for all } a, b \in R, a \neq 0 \text{ and } b \neq 0,$$

and for nonzero elements  $a$  of the ring, a function  $\varphi(a)$  is defined taking nonnegative integer values and exhibiting the following properties:

- (1)  $\varphi(ab) \geq \varphi(a)$  for all elements  $a, b \in R, a \neq 0, b \neq 0$ .
- (2) For all elements  $a, b \in R$ , where  $a \neq 0$ , there exist  $q, r \in R$  such that

$$b = aq + r \tag{13.28}$$

and either  $r = 0$  or  $\varphi(r) < \varphi(a)$ .

For the ring of integers, these properties are satisfied for  $\varphi(a) = |a|$ , while for the ring of polynomials, they are satisfied for  $\varphi(a)$  equal to the degree of the polynomial  $a$ .

**Definition 13.30** An element  $a$  of a ring  $R$  is called a *unit* or *reversible element* if there exists an element  $b \in R$  such that  $ab = 1$ . An element  $b$  is called a *divisor* of the element  $a$  (one also says that  $a$  is *divisible by*  $b$  or that  $b$  *divides*  $a$ ) if there exists an element  $c$  such that  $a = bc$ .

Clearly the property of divisibility is unchanged under multiplication of  $a$  or  $b$  by a unit. Two elements that differ by a unit are called *associates*. For example, in the ring of integers, the units are  $+1$  and  $-1$ , and associates are integers that are either equal or differ by a sign. In the ring of polynomials, the units are the constant polynomials other than the one that is identically zero, and associates are polynomials that differ from each other by a constant nonzero multiple.

An element  $p$  of a ring is *prime* if it is not a unit and has no divisors other than its associates and units.

The theory of decomposition into prime factors in a Euclidean ring repeats exactly what is known for the ring of integers.

If an element  $a$  is not prime, then it has a divisor  $b$  such that  $a = bc$ , with  $c$  not a unit. This means that  $a$  is not a divisor of  $b$ , and there exists the representation  $b = aq + r$  with  $\varphi(r) < \varphi(a)$ . But  $r = b - aq = b(1 - cq)$ , and therefore  $\varphi(r) \geq \varphi(b)$ , that is,  $\varphi(b) \leq \varphi(r) < \varphi(a)$ , which means that  $\varphi(b) < \varphi(a)$ . Applying the same reasoning to  $b$ , we finally arrive at a prime divisor  $a$ , and we shall show that every element can be represented as the product of primes. The same argument as used in the case of integers or polynomials shows the uniqueness of this decomposition in the following precise sense.

**Theorem 13.31** *If some element  $a$  in a Euclidean ring  $R$  has two factorizations into prime factors,*

$$a = p_1 \cdots p_r, \quad a = q_1 \cdots q_s,$$

*then  $r = s$ , and with a suitable numeration of the factors,  $p_i$  and  $q_i$  are associates for all  $i$ .*

As in the ring of integers, in every Euclidean ring, each element  $a \neq 0$  that is not a unit can be written in the form

$$a = up_1^{n_1} \cdots p_r^{n_r},$$

where  $u$  is a unit, all the  $p_i$  are prime elements with no two of them associates, and  $n_i$  are natural numbers. Such a representation is unique in a natural sense.

As in the ring of integers or of polynomials in one variable, representation (13.28) for  $r \neq 0$  can be applied to elements  $b$  and  $r$  and repeated until we arrive at  $r = 0$ . We will thus obtain a *greatest common divisor* (gcd) of the elements  $a$  and  $b$ , that is, a common divisor such that every other common divisor is a divisor of it. The greatest common divisor of  $a$  and  $b$  is denoted by  $d = (a, b)$  or  $d = \gcd(a, b)$ . This process, as it is for integers, is called the *Euclidean algorithm* (whence the name *Euclidean ring*). It follows from the Euclidean algorithm that a greatest common divisor of elements  $a$  and  $b$  can be written in the form  $d = ax + by$ , where  $x$  and  $y$  are some elements of the ring  $R$ .

Two elements  $a$  and  $b$  are said to be *relatively prime* if their only common divisors are units. Then we may consider that  $\gcd(a, b) = 1$ , and as follows from the Euclidean algorithm, there exist elements  $x, y \in R$  such that

$$ax + by = 1. \tag{13.29}$$

Let us now recall that the theorem on Jordan normal form holds in the case of *finite-dimensional* vector spaces, and that the fundamental theorem of abelian groups holds for *finite* abelian groups. Let us now derive analogous finiteness conditions for modules.

**Definition 13.32** A module  $M$  is said to be *finitely generated* if it contains a finite collection of elements  $\mathbf{m}_1, \dots, \mathbf{m}_r$ , called *generators*, such that every element  $\mathbf{m} \in M$  can be expressed in the form

$$\mathbf{m} = a_1\mathbf{m}_1 + \dots + a_r\mathbf{m}_r \quad (13.30)$$

for some elements  $a_1, \dots, a_r$  of the ring  $R$ .

For a vector space considered as a module over a certain field, this is the definition of finite dimensionality, and representation (13.30) is a representation of a vector  $\mathbf{m}$  in the form of a linear combination of vectors  $\mathbf{m}_1, \dots, \mathbf{m}_r$  (let us note that the system of vectors  $\mathbf{m}_1, \dots, \mathbf{m}_r$  will in general not be a basis, since we did not introduce the concept of linear independence). In the case of a finite abelian group, we may generally take for  $\mathbf{m}_1, \dots, \mathbf{m}_r$ , all the elements of the group.

Let us formulate one additional condition of the same type.

**Definition 13.33** An element  $\mathbf{m}$  of a module  $M$  over a ring  $R$  is said to be a *torsion element* if there exists an element  $a_m \neq 0$  of the ring  $R$  such that

$$a_m\mathbf{m} = \mathbf{0},$$

where  $\mathbf{0}$  is the null element of the module  $M$ , and the subscript in  $a_m$  is introduced to show that this element depends on  $\mathbf{m}$ . A module is called a *torsion module* if all of its elements are torsion elements.

In a finitely generated torsion module, there is an element  $a \neq 0$  of the ring  $R$  such that  $a\mathbf{m} = \mathbf{0}$  for all elements  $\mathbf{m} \in M$ . Indeed, it suffices to set  $a = a_{\mathbf{m}_1} \cdots a_{\mathbf{m}_r}$  for the elements  $\mathbf{m}_1, \dots, \mathbf{m}_r$  in representation (13.30). If the ring  $R$  is Euclidean, then we can conclude that  $a \neq 0$ . For the case of a finite abelian group, we may take  $a$  to be the order of the group.

*Example 13.34* Let  $M$  be a module determined by a vector space  $L$  of dimension  $n$  and by a linear transformation  $\mathcal{A}$  according to formula (13.27). For an arbitrary vector  $\mathbf{e} \in L$ , let us consider the vectors

$$\mathbf{e}, \quad \mathcal{A}(\mathbf{e}), \quad \dots, \quad \mathcal{A}^n(\mathbf{e}).$$

Their number,  $n + 1$ , is greater than the dimension  $n$  of the space  $L$ , and therefore, these vectors are linearly dependent, which means that there exists a polynomial  $f(x)$ , not identically zero, such that  $f(\mathcal{A})(\mathbf{e}) = \mathbf{0}$ , that is, in our module  $M$ , the element  $\mathbf{e}$  is a torsion element.

But if, as we did in Example 13.27, we view a vector space as a module over the field  $\mathbb{R}$  or  $\mathbb{C}$ , then not a single nonnull vector will be a torsion element of the module.

Let  $M$  be a module over a ring  $R$ . A subgroup  $M'$  of the group  $M$  is called a *submodule* if for all elements  $a \in R$  and  $\mathbf{m}' \in M'$ , we have  $a\mathbf{m}' \in M'$ .

**Example 13.35** It is obvious that every subgroup of an abelian group viewed as a module over the ring of integers is a submodule. Analogously, for a vector space viewed as a module over a ring coinciding with a suitable field, every subspace is a submodule. If  $M$  is a module defined by a vector space  $L$  and a linear transformation  $\mathcal{A}$  of  $L$  according to formula (13.27), then as is easily verified, every submodule of  $M$  is a vector subspace that is invariant with respect to the transformation  $\mathcal{A}$ .

If  $M' \subset M$  is a submodule, and  $\mathbf{m}$  is any element of the module  $M$ , then it is easily verified that the collection of all elements of the form  $a\mathbf{m} + \mathbf{m}'$ , where  $a$  is an arbitrary element of the ring  $R$ , and  $\mathbf{m}'$  is an arbitrary element of the submodule  $M'$ , is a submodule. We shall denote it by  $(\mathbf{m}, M')$ .

Since we are assuming that the ring  $R$  is Euclidean, it follows that for every torsion element  $\mathbf{m} \in M$ , there exists an element  $a \in R$  that exhibits the property  $a\mathbf{m} = \mathbf{0}$  and is such that  $\varphi(a)$  is the smallest value among all elements with this property. Then every element  $c$  for which  $c\mathbf{m} = \mathbf{0}$  is divisible by  $a$ . Indeed, if such were not the case, we would have the relationship

$$c = aq + r, \quad \varphi(r) < \varphi(a),$$

and clearly  $r\mathbf{m} = \mathbf{0}$ , which contradicts the definition of  $a$ . In particular, two such elements  $a$  and  $a'$  divide each other; that is, they are associates. The element  $a \in R$  is called the *order* of the element  $\mathbf{m} \in M$ . One must keep in mind that this expression is not quite precise, since order is defined only up to associates.

**Example 13.36** If, as in Example 13.28, a module is a vector space  $L$  viewed as a module over the polynomial ring  $f(x)$  with the aid of formula (13.27), then every element  $\mathbf{e} \in L$  is a torsion element, and its order is the same as the minimal polynomial of the vector  $\mathbf{e}$  (see the definition on p. 146), and the indicated property (every element  $c$  for which  $c\mathbf{m} = \mathbf{0}$  is divisible by the order of the element  $\mathbf{m}$ ) coincides with Theorem 4.23.

**Definition 13.37** A submodule  $M'$  of a module  $M$  is said to be *cyclic* if it contains an element  $\mathbf{m}'$  such that all the elements of the module  $M'$  can be represented in the form  $a\mathbf{m}'$  with some  $a \in R$ . This is written  $M' = \{\mathbf{m}'\}$ .

**Definition 13.38** A module  $M$  is called the *direct sum* of its submodules  $M_1, \dots, M_r$  if every element  $\mathbf{m} \in M$  can be written as a sum

$$\mathbf{m} = \mathbf{m}_1 + \dots + \mathbf{m}_r, \quad \mathbf{m}_i \in M_i,$$

and such a representation is unique. It is obvious that to establish the uniqueness of this decomposition, it suffices to prove that if  $\mathbf{m}_1 + \dots + \mathbf{m}_r = \mathbf{0}$ ,  $\mathbf{m}_i \in M_i$ , then  $\mathbf{m}_i = \mathbf{0}$  for all  $i$ . This can be written as the equality

$$M = M_1 \oplus \dots \oplus M_r.$$

The fundamental theorem that we shall prove, which contains Theorem 5.12 on the Jordan normal form and Theorem 13.22 on finite abelian groups as special cases, is the following.

**Theorem 13.39** *Every finitely generated torsion module  $M$  over a Euclidean ring  $R$  is the direct sum of cyclic submodules*

$$M = C_1 \oplus \cdots \oplus C_r, \quad C_i = \langle \mathbf{m}_i \rangle, \quad (13.31)$$

such that the order of each element  $\mathbf{m}_i$  is a power of a prime element of the ring  $R$ .

**Example 13.40** If  $M$  is a finite abelian group viewed as a module over the ring of integers, then this theorem reduces directly to the fundamental theorem of finite abelian groups (Theorem 13.22).

Let the module  $M$  be determined by the finite-dimensional complex vector space  $L$  and the linear transformation  $\mathcal{A}$  of  $L$  according to formula (13.27). Then the  $C_i$  are vector subspaces invariant with respect to  $\mathcal{A}$ , and in each of these, there exists a vector  $\mathbf{m}_i$  such that all the remaining vectors can be written in the form  $f(\mathcal{A})(\mathbf{m}_i)$ . The prime elements in the ring of complex polynomials are the polynomials of the form  $x - \lambda$ . By assumption, for each vector  $\mathbf{m}_i$ , there exist some  $\lambda_i$  and a natural number  $n_i$  such that

$$(\mathcal{A} - \lambda_i \mathcal{E})^{n_i}(\mathbf{m}_i) = \mathbf{0}.$$

If we take the smallest possible value  $n_i$ , then as proved in Sect. 5.1, the vectors

$$\mathbf{m}_i, \quad (\mathcal{A} - \lambda_i \mathcal{E})(\mathbf{m}_i), \quad \dots, \quad (\mathcal{A} - \lambda_i \mathcal{E})^{n_i-1}(\mathbf{m}_i)$$

will form a basis of this subspace, that is,  $C_i$  is a cyclic subspace corresponding to the principal vector  $\mathbf{m}_i$ . We obtain the fundamental theorem on Jordan form (Theorem 5.12).

Let us recall that we proved Theorem 5.12 by induction on the dimension of the space. More precisely, for a linear transformation  $\mathcal{A}$  on the space  $L$ , we constructed a subspace  $L'$  invariant with respect to  $\mathcal{A}$  of dimension 1 less and proved the theorem for  $L$  on the assumption that it had been proved already for  $L'$ . In fact, this meant that we constructed a sequence of nested subspaces

$$L = L_0 \supset L_1 \supset L_2 \supset \cdots \supset L_n \supset L_{n+1} = \{\mathbf{0}\}, \quad (13.32)$$

invariant with respect to  $\mathcal{A}$  and such that  $\dim L_{i+1} = \dim L_i - 1$ . Then we reduced the proof of Theorem 5.12 for  $L$  to the proof of the theorem for  $L_1$ , then for  $L_2$ , and so on. Now our first goal will be to construct in every finitely generated torsion module a sequence of submodules analogous to the sequence of subspaces (13.32).

**Lemma 13.41** *In every finitely generated torsion module  $M$  over a Euclidean ring  $R$ , there exists a sequence of submodules*

$$M = M_0 \supset M_1 \supset M_2 \supset \cdots \supset M_n \supset M_{n+1} = \{\mathbf{0}\} \quad (13.33)$$

such that  $M_i \neq M_{i+1}$ ,  $M_i = (\mathbf{m}_i, M_{i+1})$ , where  $\mathbf{m}_i$  are elements of the module  $M$ , and for each of these, there exists a prime element  $p_i$  of the ring  $R$  such that  $p_i \mathbf{m}_i \in M_{i+1}$ .

*Proof* By the definition of a finitely generated module, there exists a finite number of generators  $\mathbf{m}_1, \dots, \mathbf{m}_r \in M$  such that the elements  $a_1 \mathbf{m}_1 + \dots + a_r \mathbf{m}_r$  exhaust all the elements of the module  $M$  as  $a_1, \dots, a_r$  run through all elements of the ring  $R$ . The collection of elements of the form  $a_k \mathbf{m}_k + \dots + a_r \mathbf{m}_r$ , where  $a_k, \dots, a_r$  are all possible elements of the ring  $R$ , obviously forms a submodule of the module  $M$ . Let us denote it by  $\overline{M}_k$ . It is obvious that  $\overline{M}_k \supset \overline{M}_{k+1}$  and  $\overline{M}_k = (\mathbf{m}_k, \overline{M}_{k+1})$ . Without loss of generality, we may assume that  $\mathbf{m}_k \notin \overline{M}_{k+1}$ , since otherwise, the element  $\mathbf{m}_k$  can be excluded from among the generators. The constructed chain of submodules  $\overline{M}_k$  is still not the chain of submodules  $M_i$  that figures in Lemma 13.16. We obtain that chain from the chain of submodules  $\overline{M}_k$  by putting several intermediate submodules between the modules  $\overline{M}_k$  and  $\overline{M}_{k+1}$ .

Since  $\mathbf{m}_k \in M$  is a torsion element, there exists an element  $a \in R$  for which  $a \mathbf{m}_k = \mathbf{0}$  and in particular,  $a \mathbf{m}_k \in \overline{M}_{k+1}$ . Let  $\overline{a}$  be an element of the ring  $R$  for which  $\overline{a} \mathbf{m}_k \in \overline{M}_{k+1}$  and  $\varphi(\overline{a})$  assumes the smallest value among elements with this property. If the element  $\overline{a}$  is prime, then we set  $p_i = \overline{a}$ , and then it is unnecessary to place a submodule between  $\overline{M}_k$  and  $\overline{M}_{k+1}$ . But if  $\overline{a}$  is not prime, then let  $p_1$  be one of its prime divisors and  $\overline{a} = p_1 \overline{b}$ . Let us set  $\mathbf{m}_{k,1} = \overline{b} \mathbf{m}_k$  and  $\overline{M}_{k,1} = (\mathbf{m}_{k,1}, \overline{M}_{k+1})$ . Then clearly,  $p_1 \mathbf{m}_{k,1} \in \overline{M}_{k,1}$  and  $\overline{b} \mathbf{m}_k \in \overline{M}_{k,1}$ . As we have seen,  $\varphi(\overline{b}) < \varphi(\overline{a})$  (strict inequality). Therefore, repeating this process a finite number of times, we will place a finite number of submodules (13.33) with the required properties between  $\overline{M}_k$  and  $\overline{M}_{k+1}$ .  $\square$

*Remark 13.42* It is possible to show that the length of every chain of the form (13.33) satisfying the conditions of Lemma 13.16 is the same number  $n$ . Moreover, every chain of submodules

$$M = M_0 \supset M_1 \supset M_2 \supset \dots \supset M_m$$

in which  $M_i \neq M_{i+1}$  has length  $m \leq n$ , and this holds with much milder restrictions on the ring  $R$  and module  $M$  than we have assumed in this chapter. What is of essence here is only that between any two neighboring submodules  $M_i$  and  $M_{i+1}$ , there does not exist an “intermediate” submodule  $M'_i$  different from  $M_i$  and  $M_{i+1}$  such that  $M_i \supset M'_i \supset M_{i+1}$ .

For example, let us consider an  $n$ -dimensional vector space  $L$  over a field  $\mathbb{K}$  as a module over the ring  $R = \mathbb{K}$ . Let  $\mathbf{a}_1, \dots, \mathbf{a}_n$  be some basis. Then the subspaces  $L_i = \langle \mathbf{a}_i, \dots, \mathbf{a}_n \rangle$ ,  $i = 1, \dots, n$ , have the indicated property. Using this, we could give a definition of the dimension of a vector space without appealing to the notion of linear dependence. Thus the length  $n$  of all chains of the form (13.33) satisfying the conditions of Lemma 13.16 is the “correct” generalization of dimension of a space to finitely generated torsion modules.

The following lemma is analogous to the one we used in the proof of Theorems 5.12 and 13.22.

**Lemma 13.43** *If the order of an element  $\mathbf{m}$  of a module  $M$  is the power of a prime element,  $p^n \mathbf{m} = \mathbf{0}$ , and an element  $\mathbf{x}$  of the cyclic submodule  $\{\mathbf{m}\}$  is not divisible by  $p$  (that is, not representable in the form  $\mathbf{x} = p\mathbf{y}$ , where  $\mathbf{y} \in M$ ), then  $\{\mathbf{m}\} = \{\mathbf{x}\}$ .*

*Proof* It is obvious that  $\{\mathbf{x}\} \subset \{\mathbf{m}\}$ . Thus it remains to show that  $\{\mathbf{m}\} \subset \{\mathbf{x}\}$ , and for this, it suffices to ascertain that  $\mathbf{m} \in \{\mathbf{x}\}$ . By assumption,  $\mathbf{x} = a\mathbf{m}$ , where  $a$  is some element of the ring  $R$ . If  $a$  is divisible by  $p$ , then clearly,  $\mathbf{x}$  is also divisible by  $p$ . Indeed, if  $a = pb$  with some  $b \in R$ , then from the equality  $\mathbf{x} = a\mathbf{m}$ , we obtain  $\mathbf{x} = p\mathbf{y}$ , where  $\mathbf{y} = b\mathbf{m}$ , contradicting the assumption that  $\mathbf{x}$  is not divisible by  $p$ .

This means that  $a$  and  $p$  are relatively prime, and consequently, in view of the uniqueness of the decomposition into prime elements of the ring  $R$ ,  $a$  is also relatively prime to  $p^n$ . Then on the basis of the Euclidean algorithm, we can find elements  $u$  and  $v$  in  $R$  such that  $au + p^n v = 1$ . Multiplying both sides of this equality by  $\mathbf{m}$ , we obtain that  $\mathbf{m} = u\mathbf{x}$ , which means that  $\mathbf{m} \in \{\mathbf{x}\}$ .  $\square$

**Lemma 13.44** *Let  $M_1$  be a submodule of the module  $M$  over a Euclidean ring  $R$  such that  $M = (\mathbf{m}, M_1)$  and  $M \neq M_1$ . Then if for some  $a, p \in R$ , we have the inclusions  $a\mathbf{m} \in M_1$  and  $p\mathbf{m} \in M_1$ , where the element  $p$  is prime, then  $a$  is divisible by  $p$ .*

*Proof* Let us assume that  $a$  is not divisible by  $p$ . Since the element  $p$  is prime, we have  $(a, p) = 1$ , and from the Euclidean algorithm in the ring  $R$ , it follows that there exist two elements  $u, v \in R$  for which  $au + pv = 1$ . Multiplying both sides of this equality by  $\mathbf{m}$ , taking into account the inclusions  $a\mathbf{m} \in M_1$  and  $p\mathbf{m} \in M_1$ , we obtain that  $\mathbf{m} \in M_1$ . By definition,  $(\mathbf{m}, M_1)$  consists of elements  $b\mathbf{m} + \mathbf{m}'$  for all possible  $b \in R$  and  $\mathbf{m}' \in M_1$ . Therefore,  $M = (\mathbf{m}, M_1) = M_1$ , which contradicts the assumption of the lemma.  $\square$

*Proof of Theorem 13.39* The proof is an almost verbatim repetition of the proof of Theorems 5.12 and 13.22. We may use induction on the length  $n$  of the chain (13.33), that is, we may assume the theorem to be true for the module  $M_1$ . Let

$$M_1 = C_1 \oplus \cdots \oplus C_r, \quad (13.34)$$

where  $C_i = \{c_i\}$  are cyclic submodules, and the order of each element  $c_i$  is the power of a prime element. By Lemma 13.16,  $M = (\mathbf{m}, M_1)$  and  $p\mathbf{m} \in M_1$ , where  $p$  is a prime element. Then based on the decomposition (13.34), we have

$$p\mathbf{m} = z_1 + \cdots + z_r, \quad z_i \in C_i. \quad (13.35)$$

We shall select those elements  $z_i$  that are divisible by  $p$ . By a change in numeration, we may assume that these are the first  $s - 1$  terms. Let us set  $z_i = pz'_i$  for  $i = 1, \dots, s - 1$ . We must now consider two cases.



*Case 1:* The number  $s - 1$  is equal to  $r$ . Then  $p\mathbf{m} = p\mathbf{m}'$ , where  $\mathbf{m}' = z'_1 + \cdots + z'_r$ . Let us set  $\mathbf{m} - \mathbf{m}' = \overline{\mathbf{m}}$ . It is obvious that  $p\overline{\mathbf{m}} = \mathbf{0}$ . We shall prove that the module  $M$  can be written in the form

$$M = \{\overline{\mathbf{m}}\} \oplus C_1 \oplus \cdots \oplus C_r.$$

Indeed, by assumption, every element  $\mathbf{x} \in M$  can be represented in the form  $\mathbf{x} = a\mathbf{m} + \mathbf{y}$ , where  $a \in R$  and  $\mathbf{y} \in M_1$ , which means also in the form  $\mathbf{x} = a\overline{\mathbf{m}} + \mathbf{y}'$ , where  $\mathbf{y}' = a\mathbf{m}' + \mathbf{y} \in M_1$ .

Let us prove that for two such representations

$$\mathbf{x} = a\overline{\mathbf{m}} + \mathbf{y}, \quad \mathbf{x} = a'\overline{\mathbf{m}} + \mathbf{y}', \quad (13.36)$$

we have the equalities  $a\overline{\mathbf{m}} = a'\overline{\mathbf{m}}$  and  $\mathbf{y} = \mathbf{y}'$ . From this it will follow that

$$M = \{\overline{\mathbf{m}}\} \oplus M_1 = \{\overline{\mathbf{m}}\} \oplus C_1 \oplus \cdots \oplus C_r,$$

which in our case, is relationship (13.31).

We obtain from equalities (13.36) that  $a\overline{\mathbf{m}} = a'\overline{\mathbf{m}}$ , where  $\overline{a} = a - a'$ ,  $\overline{\mathbf{y}} = \mathbf{y}' - \mathbf{y}$ , and by assumption,  $\overline{\mathbf{y}} \in M_1$ . By Lemma 13.16, there exists a prime element  $p$  of the ring  $R$  such that  $p\overline{\mathbf{m}} \in M_1$ , and this means that  $p\overline{\mathbf{m}} \in M_1$ . By Lemma 13.20, from the inclusions  $a\overline{\mathbf{m}} \in M_1$  and  $p\overline{\mathbf{m}} \in M_1$ , it follows that the element  $\overline{a}$  is divisible by  $p$ , that is,  $\overline{a} = bp$  for some  $b \in R$ . From this, we obviously obtain that  $a\overline{\mathbf{m}} = b(p\overline{\mathbf{m}}) = \mathbf{0}$ . Consequently,  $a\overline{\mathbf{m}} = a'\overline{\mathbf{m}}$  and  $\mathbf{y} = \mathbf{y}'$ .

*Case 2:* The number  $s - 1$  is less than  $r$ . If an element  $\mathbf{c}_i$  has order  $p_i^{n_i}$  and  $p_i$  is not an associate of  $p$ , then  $p_i^{n_i}$  is not divisible by  $p$ , and therefore, every element of the module  $C_i = \{\mathbf{c}_i\}$  is divisible  $p$ , by Lemma 13.17. Therefore, among the chosen  $s - 1$  submodules  $C_i$  are all those such that the order of the element  $\mathbf{c}_i$  is  $p_i^{n_i}$ , and  $p_i$  is not an associate of  $p$ . Since the order of an element is in general defined only up to replacing it by an associate, we may consider that in the remaining submodules  $C_s = \{\mathbf{c}_s\}, \dots, C_r = \{\mathbf{c}_r\}$ , the order of the element  $\mathbf{c}_i$  is a power of  $p$ .

By construction, in the decomposition (13.35), we have  $\mathbf{z}_i = p\mathbf{z}'_i$ ,  $\mathbf{z}'_i \in C_i$ , for all  $i = 1, \dots, s - 1$ . Setting  $\mathbf{z}'_1 + \cdots + \mathbf{z}'_{s-1} = \mathbf{z}'$  and  $\mathbf{m} - \mathbf{z}' = \overline{\mathbf{m}}$ , we obtain the equality

$$p\overline{\mathbf{m}} = \mathbf{z}_s + \cdots + \mathbf{z}_r. \quad (13.37)$$

Since the order of the element  $\mathbf{c}_i$  for  $i = s, \dots, r$  is a power of  $p$ , the order of an arbitrary element  $\mathbf{z}_i$  in the decomposition (13.37) is also a power of  $p$ . Let us denote it by  $p^{n_i}$ . Obviously, we may choose the numeration of the terms in formula (13.37) in such a way that the numbers  $n_i$  do not decrease:  $1 \leq n_s \leq n_{s+1} \leq \cdots \leq n_r$ . Let us prove that the order of the element  $\overline{\mathbf{m}}$  is equal to  $p^{n_r+1}$  and that we have the equality

$$M = \{\overline{\mathbf{m}}\} \oplus C_1 \oplus \cdots \oplus C_{s-1} \oplus \cdots \oplus C_{r-1},$$

that is, in the decomposition, all submodules  $C_i$  occur other than  $C_r$ . With this, relationship (13.31) will be proved in the second case as well; that is, the proof of Theorem 13.39 will be complete.

Multiplying both sides of equality (13.37) by  $p^{n_r}$  and using the fact that  $p^{n_r} z_i = \mathbf{0}$  for all  $i = s, \dots, r$ , we obtain that  $p^{n_r+1} \bar{m} = \mathbf{0}$ . If the order  $a$  of an element  $\bar{m}$  is not an associate of  $p^{n_r+1}$ , then it divides it, and is equal, up to an associate, to  $p^k$  for some  $k < n_r + 1$ . Multiplying relationship (13.37) by  $p^{k-1}$  and using the fact that the submodules  $C_1, \dots, C_r$  form a direct sum, we obtain that  $p^{k-1} z_i = \mathbf{0}$  for all  $i = s, \dots, r$ . In particular,  $p^{k-1} z_r = \mathbf{0}$ , and this contradicts the assumption  $k < n_r + 1$  and that the order of the element  $z_r$  is equal to  $p^{n_r}$ . Thus the order of the element  $\bar{m}$  is equal to  $p^{n_r+1}$ .

Let us note that by construction, in the decomposition (13.37), the element  $z_r$  is not divisible by  $p$ .

From what we have proved, on the basis of Lemma 13.17, it follows that  $\{z_r\} = \{c_r\} = C_r$ . From this it follows that every element  $m \in M$  can be represented as a sum of elements of the modules

$$\{\bar{m}\}, C_1, \dots, C_{s-1}, \dots, C_{r-1}. \quad (13.38)$$

Indeed, an analogous assertion holds for the modules

$$\{\bar{m}\}, C_1, \dots, C_{s-1}, \dots, C_r, \quad (13.39)$$

since by our construction,  $\bar{m} = m - z'$  and  $z' = z'_1 + \dots + z'_{s-1}$ , where  $z'_i \in C_i$ . Consequently,  $m = \bar{m} + z'_1 + \dots + z'_{s-1}$ , which means that every element  $m \in M$  is a sum of elements of the modules (13.39).

We now must verify that every element of the submodule  $C_r$  can be represented as a sum of elements of the submodules (13.38). Since  $C_r = \{z_r\}$ , it suffices to verify this for a single element  $z_r$ . But relationship (13.37) gives us precisely the required representation:

$$z_r = p\bar{m} - z_s - \dots - z_{r-1}.$$

It remains to verify the second condition entering into the definition of a direct sum: that such a representation is unique. To this end, it suffices to prove that in the relationship

$$a\bar{m} + f_1 + \dots + f_{s-1} + \dots + f_{r-1} = \mathbf{0}, \quad f_i \in C_i, \quad (13.40)$$

all the terms must equal  $\mathbf{0}$ .

Indeed, from relationship (13.40), taking into account (13.34), it follows that  $a\bar{m} \in M_1$ . But by the construction of the element  $\bar{m}$ , we then also have  $am \in M_1$ . By Lemma 13.20, from the inclusions  $am \in M_1$  and  $pm \in M_1$ , we have that the element  $a$  is divisible by  $p$ , that is,  $a = bp$  for some  $b \in R$ . Furthermore, we know that

$$p\bar{m} = z_s + \dots + z_r,$$

and moreover, the order of the element  $z_r$  is  $p^{n_r}$ , while the order of the element  $\bar{m}$  is  $p^{n_r+1}$ . On substituting all these relationships into decomposition (13.40), we obtain

$$b(z_s + \dots + z_r) + f_1 + \dots + f_{s-1} + \dots + f_{r-1} = \mathbf{0}.$$

Then it follows from formula (13.34) that  $bz_r = \mathbf{0}$ , and since the order of the element  $z_r$  is equal to  $p^{n_r}$ , we have that  $p^{n_r}$  divides  $b$ . This means that the element  $a$  is divisible by  $p^{n_r+1}$ , and  $a\overline{m} = \mathbf{0}$ . But then from equality (13.40), it follows that  $f_1 + \cdots + f_{r-1} = \mathbf{0}$ . Using again the induction hypothesis (13.34), we obtain that  $f_1 = \mathbf{0}, \dots, f_{r-1} = \mathbf{0}$ . This completes the proof of Theorem 13.39.  $\square$

For Theorem 13.39, we have the same uniqueness theorem as in the case of Theorem 5.12 and Theorem 13.22. Namely, if

$$M = C_1 \oplus \cdots \oplus C_r, \quad C_i = \{\mathbf{m}_i\}, \quad M = D_1 \oplus \cdots \oplus D_s, \quad D_j = \{\mathbf{n}_j\}$$

are two decompositions of finitely generated torsion modules  $M$  in which the orders of elements  $\mathbf{m}_i$  and  $\mathbf{n}_j$  are prime powers, that is,  $p_i^{r_i} \mathbf{m}_i = \mathbf{0}$  and  $q_j^{s_j} \mathbf{n}_j = \mathbf{0}$ , where  $p_i$  and  $q_j$  are prime elements, then with a suitable numeration of the terms  $C_i$  and  $D_j$ , elements  $p_i$  and  $q_i$  are associates, and  $r_i = s_i$ . However, a natural proof of this theorem would require some new concepts, and we shall not pursue this here.

# Chapter 14

## Elements of Representation Theory

Representation theory is one of the most “applied” branches of algebra. It has many applications in various branches of mathematics and mathematical physics. In this chapter, we shall be concerned with the problem of finding all finite-dimensional representations of finite groups. But there is an analogous theory that has been developed for certain types of infinite groups, which is important in many other branches of mathematics.

### 14.1 Basic Concepts of Representation Theory

Let us recall some definitions from the previous chapter that will play a key role here.

A homomorphism of a group  $G$  into a group  $G'$  is a mapping  $f : G \rightarrow G'$  such that for every pair of elements  $g_1, g_2 \in G$ , we have the relationship

$$f(g_1 g_2) = f(g_1) f(g_2).$$

An isomorphism of a group  $G$  onto a group  $G'$  is a bijective homomorphism  $f : G \rightarrow G'$ . Groups  $G$  and  $G'$  are said to be *isomorphic* if there exists an isomorphism  $f : G \rightarrow G'$  between them. This is denoted by  $G \simeq G'$ .

**Definition 14.1** A *representation* of a group  $G$  is a homomorphism of  $G$  into the group of nonsingular linear transformations of a vector space  $L$ . The space  $L$  is called the *space* of the representation or the *representation space*, and its dimension, that is,  $\dim L$ , is the *dimension* of the representation.

Thus in order to specify a representation of a group  $G$ , it is necessary to associate with each element  $g \in G$  a nonsingular linear transformation  $\mathcal{A}_g : L \rightarrow L$  such that for  $g_1, g_2 \in G$ , the condition

$$\mathcal{A}_{g_1 g_2} = \mathcal{A}_{g_1} \mathcal{A}_{g_2} \tag{14.1}$$

is satisfied. Since the group of nonsingular linear transformations of an  $n$ -dimensional vector space is isomorphic to the group of nonsingular square matrices of order  $n$ , to give a representation, it suffices to associate with each element  $g \in G$  a nonsingular square matrix  $\mathcal{A}_g$  such that (14.1) is satisfied.

It follows at once from (14.1) that for a representation  $\mathcal{A}_g$  and any number of elements  $g_1, \dots, g_k$  of the group  $G$ , we have the relationship

$$\mathcal{A}_{g_1 \dots g_k} = \mathcal{A}_{g_1} \cdots \mathcal{A}_{g_k}. \quad (14.2)$$

Moreover, it is obvious that if  $e$  is the identity element of  $G$ , then

$$\mathcal{A}_e = \mathcal{E}, \quad (14.3)$$

where  $\mathcal{E}$  is the identity linear transformation of the space  $L$ . And if  $g^{-1}$  is the inverse of the element  $g$ , then

$$\mathcal{A}_{g^{-1}} = \mathcal{A}_g^{-1}, \quad (14.4)$$

that is,  $\mathcal{A}_{g^{-1}}$  is the transformation that is the inverse of  $\mathcal{A}_g$ .

*Example 14.2* Let  $G = GL_n$  be the group of nonsingular square matrices of order  $n$ . For each matrix  $g \in GL_n$ , let us set

$$\mathcal{A}_g = |g|.$$

Since  $|g|$  is a number, which by assumption is different from zero, we have a one-dimensional representation. It is obvious that for every integer  $n$ , the equality

$$\mathcal{B}_g = |g|^n$$

will also define a one-dimensional representation.

*Example 14.3* Let  $G = S_n$  be the symmetric group of degree  $n$ , that is, the group of permutations of an  $n$ -element set  $M$ , and let  $L$  be a vector space of dimension  $n$ , in which we have chosen a basis  $e_1, \dots, e_n$ . For the representation

$$g = \begin{pmatrix} 1 & 2 & \cdots & n \\ j_1 & j_2 & \cdots & j_n \end{pmatrix},$$

let us define  $\mathcal{A}_g$  as the linear transformation such that

$$\mathcal{A}_g(e_1) = e_{j_1}, \quad \mathcal{A}_g(e_2) = e_{j_2}, \quad \dots, \quad \mathcal{A}_g(e_n) = e_{j_n}.$$

Then we obtain an  $n$ -dimensional representation of the group  $S_n$ .

To avoid having to use a specific numeration of the elements of the set  $M$ , let us associate with the element  $a \in M$ , the basis vector  $e_a$ . Then the representation described above is given by the formula

$$\mathcal{A}_g(e_a) = e_b \quad \text{if } g(a) = b,$$

for every transformation  $g : M \rightarrow M$ .

*Example 14.4* Let  $G = S_3$  be the symmetric group of degree 3, and let  $L$  be a two-dimensional space with basis  $e_1, e_2$ . Let us define a vector  $e_3$  by  $e_3 = -(e_1 + e_2)$ . For the representation

$$g = \begin{pmatrix} 1 & 2 & 3 \\ j_1 & j_2 & j_3 \end{pmatrix},$$

let us define  $\mathcal{A}_g$  as the transformation such that

$$\mathcal{A}_g(e_1) = e_{j_1}, \quad \mathcal{A}_g(e_2) = e_{j_2}.$$

It is easily verified that in this way, we obtain a two-dimensional representation of the symmetric group  $S_3$ .

*Example 14.5* Let  $G = GL_2$  be the group of nonsingular matrices of order 2, and let  $L$  be the space of polynomials in the two variables  $x$  and  $y$  whose total degree in both variables does not exceed  $n$ . For a nonsingular matrix

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

let us define  $\mathcal{A}_g$  as the linear transformation of the space  $L$  taking polynomials  $f(x, y)$  to  $f(ax + by, cx + dy)$ , that is,

$$\mathcal{A}_g(f(x, y)) = f(ax + by, cx + dy).$$

It is easy to verify that relationship (14.1) is satisfied in this case, that is, we have a representation of the group of nonsingular matrices of order 2. Its dimension is equal to the dimension of the space of polynomials in  $x$  and  $y$  whose dimension (in both variables combined) does not exceed  $n$ ; that is, as is easily seen, it is equal to  $(n+1)(n+2)/2$ .

*Example 14.6* For any group and an  $n$ -dimensional space  $L$ , the representation defined by the formula  $\mathcal{A}_g = \mathcal{E}$ , where  $\mathcal{E}$  is the identity transformation on the space  $L$ , is called the  *$n$ -dimensional identity representation*.

In the definition of a representation, the space  $L$  can also be *infinite-dimensional*. In this case, the representation is also said to be *infinite-dimensional*. For example, defining a representation just as in Example 14.5, but taking for  $L$  the space of all continuous functions, we obtain an infinite-dimensional representation. In the sequel, we shall consider only finite-dimensional representations, and we shall always consider the space  $L$  to be complex.

*Example 14.7* Representations of the symmetric group  $S_n$  are of interest in many problems. All such representations are known, but we shall describe here only the one-dimensional representations of the group  $S_n$ . In this case, a nonsingular linear transformation  $\mathcal{A}_g$  is given by a matrix of order 1, that is, a single complex number (which, of course, is nonzero). We thereby arrive at a function on the group taking

numeric values. Let us denote this function by  $\varphi(g)$ . Then by definition, it must satisfy the conditions  $\varphi(g) \neq 0$  and

$$\varphi(gh) = \varphi(g)\varphi(h) \quad (14.5)$$

for all elements  $g$  and  $h$  in the group  $S_n$ .

It is easy to find all possible values  $\varphi(\tau)$  if  $\tau$  is a transposition. Namely, setting  $g = h = \tau$  and using the facts that  $\tau^2 = e$  (the identity transformation) and that obviously,  $\varphi(e) = 1$ , we obtain from relationship (14.5) the equality  $\varphi(\tau)^2 = 1$ , from which follows  $\varphi(\tau) = \pm 1$ . It is theoretically possible that for some transpositions,  $\varphi(\tau) = 1$ , while for others,  $\varphi(\tau) = -1$ . However, in reality, such is not the case, and one of the equalities  $\varphi(\tau) = 1$  and  $\varphi(\tau) = -1$  holds for all transpositions  $\tau$ , with the choice of sign depending only on the one-dimensional representation  $\varphi$ . Let us prove this.

Let  $\tau = \tau_{a,b}$  and  $\tau' = \tau_{c,d}$  be two transpositions, where  $a, b, c, d$  are elements of the set  $M$  (see formula (13.3)). Obviously, there exists a permutation  $g$  of the set  $M$  such that  $g(c) = a$  and  $g(d) = b$ . Then as is easily verified, based on the definition of a transposition, we have the equality  $g^{-1}\tau_{a,b}g = \tau_{c,d}$ , that is,  $\tau' = g^{-1}\tau g$ . In view of relationships (14.2), (14.4), and (14.5), we obtain from the last equality that

$$\varphi(\tau') = \varphi(g)^{-1}\varphi(\tau)\varphi(g) = \varphi(\tau),$$

which proves our assertion for all transpositions  $\tau$  and  $\tau'$ . We shall now make use of the fact that every element  $g$  of the group  $S_n$  is the product of a finite number of transpositions; see formula (13.4). Taking the aforesaid into account, it follows from this that

$$\varphi(g) = \varphi(\tau_{a_1,b_1})\varphi(\tau_{a_2,b_2}) \cdots \varphi(\tau_{a_k,b_k}) = \varphi(\tau)^k, \quad (14.6)$$

where  $\varphi(\tau) = +1$  or  $-1$ .

Thus there are two possible cases. The first case is that for all transpositions  $\tau \in S_n$ , the number  $\varphi(\tau)$  is equal to 1. In view of formula (14.6), for every transposition  $g \in S_n$ , we have  $\varphi(g) = 1$ , that is, the function  $\varphi$  on  $S_n$  is identically equal to 1, and therefore, it gives the one-dimensional identity representation of the group  $S_n$ . The second case is that for all transpositions  $\tau \in S_n$ , we have  $\varphi(\tau) = -1$ . Then, in view of formula (14.6), for a transposition  $g \in S_n$ , we have  $\varphi(g) = (-1)^k$ , where  $k$  corresponds to the parity of the transposition  $g$ . In other words,  $\varphi(g) = 1$  if the transposition  $g$  is even, and  $\varphi(g) = -1$  if the transposition  $g$  is odd. From relationship (13.4), it follows at once that such a function  $\varphi$  indeed determines a one-dimensional representation of the group  $S_n$ , which we denote by  $\varepsilon(g)$ .

Thus we have obtained the following result: *the symmetric group  $S_n$  has exactly two one-dimensional representations: the identity and  $\varepsilon(g)$ .*

One-dimensional representations of the group  $S_n$  and related groups (such as the alternating group  $A_n$ ) play a large role in a variety of questions in algebra. For example, one of the best-known results in algebra is the derivation of formulas for the solution of equations of degrees 3 and 4. For a long time, mathematicians were thwarted in their attempts to find analogous formulas for equations of degree 5 and

higher. Finally, it was proved that such an attempt was futile, that is, that *there exists no formula that expresses the roots of a polynomial equation of degree 5 or greater in terms of its coefficients using the usual arithmetic operations and the extraction of roots of arbitrary degree*. A key point in the proof of this assertion was the establishment of the fact that the alternating group  $A_n$  for  $n \geq 5$  has no one-dimensional representation other than the identity. For  $n = 3$  and 4, such representations of the group  $A_n$  exist, and that is what explains the existence of formulas for the solution of equations of those degrees.

Now let us establish what representations we shall consider to be identical.

**Definition 14.8** Two representations  $g \mapsto \mathcal{A}_g$  and  $g \mapsto \mathcal{A}'_g$  of the same group  $G$  with spaces  $L$  and  $L'$  of the same dimension are said to be *equivalent* if there exists an isomorphism  $\mathcal{C} : L' \rightarrow L$  of the vector spaces  $L'$  and  $L$  such that

$$\mathcal{A}'_g = \mathcal{C}^{-1} \mathcal{A}_g \mathcal{C} \quad (14.7)$$

for every element  $g \in G$ .

Let  $e'_1, \dots, e'_n$  be a basis of the space  $L'$  and let  $e_1 = \mathcal{C}(e'_1), \dots, e_n = \mathcal{C}(e'_n)$  be the corresponding basis of the space  $L$ , since the linear transformation  $\mathcal{C} : L' \rightarrow L$  is an isomorphism. Comparing relationship (14.7) with the change-of-matrix formula (3.43), we see that this definition means that the matrix of the transformation  $\mathcal{A}'_g$  with basis  $e'_1, \dots, e'_n$  coincides with the matrix of the transformation  $\mathcal{A}_g$  with basis  $e_1, \dots, e_n$ . Thus the representations  $\mathcal{A}_g$  and  $\mathcal{A}'_g$  are equivalent if and only if one can choose bases in the spaces  $L$  and  $L'$  such that for each element  $g \in G$ , the transformations  $\mathcal{A}_g : L \rightarrow L$  and  $\mathcal{A}'_g : L' \rightarrow L'$  have identical matrices.

Let  $g \mapsto \mathcal{A}_g$  be a representation of the group  $G$ , and let  $L$  be its representation space. A subspace  $M \subset L$  is said to be *invariant* with respect to the representation  $\mathcal{A}_g$  if it is invariant with respect to all linear transformations  $\mathcal{A}_g : L \rightarrow L$  for all  $g \in G$ . Let us denote by  $\mathcal{B}_g$  the restriction of  $\mathcal{A}_g$  to the subspace  $M$ . It is obvious that  $\mathcal{B}_g$  is a representation of the group  $G$  with representation space  $M$ . The representation  $\mathcal{B}_g$  is said to be the representation *induced* by the representation  $\mathcal{A}_g$  with invariant subspace  $M$ . This is also expressed by saying that the representation  $\mathcal{B}_g$  is *contained* in the representation  $\mathcal{A}_g$ .

**Example 14.9** Let us consider the  $n$ -dimensional representation of the group  $S_n$  described in Example 14.3. As is easily verified, the collection of all vectors of the form  $\sum_{a \in M} \alpha_a e_a$ , where  $\alpha_a$  is an arbitrary scalar satisfying  $\sum_{a \in M} \alpha_a = 0$ , forms a subspace  $L' \subset L$  of dimension  $n - 1$ , invariant with respect to this representation. The representation thus induced in  $L'$  is an  $(n - 1)$ -dimensional representation of the group  $S_n$ . In the case  $n = 3$ , it is equivalent to the representation of the group  $S_3$  described in Example 14.4.

**Example 14.10** In Example 14.5, let us denote by  $M_k$  ( $k = 0, \dots, n$ ) the subspace consisting of polynomials of degree at most  $k$  in the variables  $x$  and  $y$ . Each of  $M_k$  is an invariant subspace of every  $M_l$  with index  $l \geq k$ .



**Definition 14.11** A representation is said to be *reducible* if its representation space  $L$  has an invariant subspace different from  $(0)$  and from all of  $L$ . Otherwise, it is said to be *irreducible*.

Examples 14.3 and 14.5 give reducible representations. Clearly, the  $n$ -dimensional identity representation is reducible if  $n > 1$ : every subspace of the representation space is invariant. Every one-dimensional representation is irreducible.

Let us prove that the representation in Example 14.4 is irreducible. Indeed, any invariant subspace different from  $(0)$  and  $L$  must be one-dimensional. Let  $u$  be a basis vector of such a subspace. The condition of invariance means that

$$\mathcal{A}_g(u) = \lambda_g u$$

for every  $g \in S_3$ , where  $\lambda_g$  is some scalar depending on the element  $g$ , that is,  $u$  is a common eigenvector for all transformations  $\mathcal{A}_g$ . It is easy to verify that this is impossible: the eigenvectors of the transformation  $\mathcal{A}_{g_1}$  with  $g_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$  have the form  $\alpha(e_1 + e_2)$  and  $\beta(e_1 - e_2)$ , and the eigenvectors of the transformation  $\mathcal{A}_{g_2}$  with  $g_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$  have the form  $\gamma e_2$  and  $\delta(2e_1 + e_2)$ , and these clearly cannot coincide.

**Definition 14.12** A representation  $\mathcal{A}_g$  is said to be the *direct sum* of the  $r$  representations

$$\mathcal{A}_g^{(1)}, \dots, \mathcal{A}_g^{(r)}$$

if its representation space  $L$  is the direct sum of the  $r$  invariant subspaces

$$L = L_1 \oplus \dots \oplus L_r, \quad (14.8)$$

and  $\mathcal{A}_g$  induces in every  $L_i$  a representation equivalent to  $\mathcal{A}_g^{(i)}$ ,  $i = 1, \dots, r$ .

*Example 14.13* The  $n$ -dimensional identity representation is the direct sum of  $n$  one-dimensional identity representations. To convince oneself of this, it suffices to decompose the space of this representation in some way into a direct sum of one-dimensional subspaces.

*Example 14.14* In the situation of Example 14.9, let us denote by  $L_1$  an invariant subspace  $L'$  of dimension  $n - 1$ , and let us denote by  $L_2$  the one-dimensional subspace spanned by the vector  $\sum_{a \in M} e_a$ . Clearly,  $L_2$  is also an invariant subspace of this representation, and we have the decomposition  $L = L_1 \oplus L_2$ . In particular, the representation introduced in Example 14.3, for  $n = 3$ , is the direct sum of the representation of Example 14.4 and the one-dimensional identity representation.

It can happen that the representation space  $L$  has an invariant subspace  $L_1$ , yet it is impossible to find a complementary invariant subspace  $L_2$  such that  $L = L_1 \oplus L_2$ . In other words, the representation is reducible, but it is not the direct sum of two other representations.

*Example 14.15* Let  $G = \{g\}$  be an infinite cyclic group, and let  $L$  be a two-dimensional space with basis  $e_1, e_2$ . Let us denote by  $\mathcal{A}_n$  the transformation having

in this basis the matrix  $\begin{pmatrix} 1 & 0 \\ n & 1 \end{pmatrix}$ . It is obvious that  $\mathcal{A}_n \mathcal{A}_m = \mathcal{A}_{n+m}$ . From this, it follows that on setting  $\mathcal{A}_{g^n} = \mathcal{A}_n$ , we obtain a representation of the group  $G$ . The line  $L_1 = \langle e_2 \rangle$  is an invariant subspace:  $\mathcal{A}_n(e_2) = e_2$ . However, there are no other invariant subspaces. Thus, for instance, the transformation  $\mathcal{A}_1$  has no eigenvectors other than  $e_2$ . Therefore, our representation is reducible, but it is not a direct sum.

Let us note that in Example 14.15, the group  $G$  was infinite. It turns out that for finite groups, such a phenomenon cannot occur. Namely, in the following section, it will be proved that if a representation  $\mathcal{A}_g$  of a finite group is reducible, that is, the vector space  $L$  of this representation contains an invariant subspace  $L_1$ , then  $L$  is the direct sum of  $L_1$  and another invariant subspace  $L_2$ . Hence it follows that every representation of a finite group is the direct sum of irreducible representations. As regards irreducible representations, it will be proved in Sect. 14.3 that (up to equivalence) there is only of finite number of them.

From this point on, to the end of this book, we shall always assume that a group  $G$  is finite, with the sole exception of Example 14.36.

## 14.2 Representations of Finite Groups

The proof of the fundamental property of representations of finite groups formulated at the end of the preceding section uses several properties of complex vector spaces.

Let us consider a representation of a finite group  $G$ . Let  $L$  be its representation space. Let us define on  $L$  some Hermitian form  $\varphi(\mathbf{x}, \mathbf{y})$  for which the corresponding quadratic-Hermitian form  $\psi(\mathbf{x}) = \varphi(\mathbf{x}, \mathbf{x})$  is positive definite, and thus it takes positive values for all  $\mathbf{x} \neq \mathbf{0}$ . For example, if  $L = \mathbb{C}^n$ , then for vectors  $\mathbf{x}$  and  $\mathbf{y}$  with coordinates  $(x_1, \dots, x_n)$  and  $(y_1, \dots, y_n)$ , let us set

$$\varphi(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n x_i \bar{y}_i.$$

In the sequel, we shall denote  $\varphi(\mathbf{x}, \mathbf{y})$  by  $(\mathbf{x}, \mathbf{y})$  and call it a *scalar product* in the space  $L$ . The concepts and simple results that we proved in Chap. 7 for Euclidean spaces can be transferred to this case verbatim. Let us list those of them that we are now going to use:

1. The *orthogonal complement* of a subspace  $L' \subset L$  is the collection of all vectors  $\mathbf{y} \in L$  for which  $(\mathbf{x}, \mathbf{y}) = 0$  for all  $\mathbf{x} \in L'$ . The orthogonal complement of a subspace  $L'$  is itself a subspace of  $L$  and is denoted by  $(L')^\perp$ . We have the decomposition  $L = L' \oplus (L')^\perp$ .
2. A *unitary transformation* (the analogue of orthogonal transformation for the case of a complex space) is a linear transformation  $\mathcal{U} : L \rightarrow L$  such that for all vectors  $\mathbf{x}, \mathbf{y} \in L$ , we have the relationship

$$(\mathcal{U}(\mathbf{x}), \mathcal{U}(\mathbf{y})) = (\mathbf{x}, \mathbf{y}).$$

3. The complex analogue of Theorem 7.24 is this: if a subspace  $L' \subset L$  is invariant with respect to a unitary transformation  $\mathcal{U}$ , then its orthogonal complement  $(L')^\perp$  is also invariant with respect to  $\mathcal{U}$ .

**Definition 14.16** A representation  $\mathcal{U}_g$  of a group  $G$  is said to be *unitarizable* if it is possible to introduce a scalar product on its representation space  $L$  such that all transformations  $\mathcal{U}_g$  become unitary.

The property of a representation being unitarizable obviously remains true under a change to an equivalent representation.

Indeed, let  $g \mapsto \mathcal{U}_g$  be a unitarizable representation of some group  $G$  with space  $L$  and Hermitian form  $\varphi(\mathbf{x}, \mathbf{y})$ . Let us consider an arbitrary isomorphism  $\mathcal{C} : L' \rightarrow L$ . As we know, it determines an equivalent representation  $g \mapsto \mathcal{U}'_g$  of the same group with space  $L'$ . Let us show that the representation  $g \mapsto \mathcal{U}'_g$  is also unitarizable. As the scalar product in  $L'$  let us choose the form defined by the relationship

$$\psi(\mathbf{u}, \mathbf{v}) = \varphi(\mathcal{C}(\mathbf{u}), \mathcal{C}(\mathbf{v})) \quad (14.9)$$

for vectors  $\mathbf{u}, \mathbf{v} \in L'$ . It is obvious that  $\psi(\mathbf{u}, \mathbf{v})$  is a Hermitian form on  $L'$  and that  $\psi(\mathbf{u}, \mathbf{u}) > 0$  for every nonnull vector  $\mathbf{u} \in L'$ . Let us verify that the scalar product  $\psi(\mathbf{u}, \mathbf{v})$  indeed establishes the unitarizability of the representation  $g \mapsto \mathcal{U}'_g$ . Substituting the vectors  $\mathcal{U}'_g(\mathbf{u})$  and  $\mathcal{U}'_g(\mathbf{v})$  into equality (14.9), taking into account (14.7) and the unitarizability of the representation  $g \mapsto \mathcal{U}_g$ , we obtain the relationship

$$\begin{aligned} \psi(\mathcal{U}'_g(\mathbf{u}), \mathcal{U}'_g(\mathbf{v})) &= \psi(\mathcal{C}^{-1}\mathcal{U}_g\mathcal{C}(\mathbf{u}), \mathcal{C}^{-1}\mathcal{U}_g\mathcal{C}(\mathbf{v})) \\ &= \varphi(\mathcal{U}_g\mathcal{C}(\mathbf{u}), \mathcal{U}_g\mathcal{C}(\mathbf{v})) = \varphi(\mathcal{C}(\mathbf{u}), \mathcal{C}(\mathbf{v})) = \psi(\mathbf{u}, \mathbf{v}), \end{aligned}$$

which means that the representation  $g \mapsto \mathcal{U}'_g$  is unitarizable.

**Lemma 14.17** *If a space  $L$  of a unitarizable representation  $\mathcal{U}_g$  of a group  $G$  contains an invariant subspace  $L'$ , then it also contains a second invariant subspace  $L''$  such that  $L = L' \oplus L''$ .*

*Proof* Let us take as  $L''$  the orthogonal complement  $(L')^\perp$ . Then the space  $L''$  is invariant with respect to all transformations  $\mathcal{U}_g$ , and we have the decomposition  $L = L' \oplus L''$ .  $\square$

The application of this lemma to representations of finite groups is based on the following fundamental fact.

**Theorem 14.18** *Every representation  $\mathcal{A}_g$  of a finite group  $G$  is unitarizable.*

*Proof* Let us introduce a scalar product on the representation space  $L$  in such a way that all linear transformations  $\mathcal{A}_g$  become unitary. For this, let us take an arbitrary scalar product  $[\mathbf{x}, \mathbf{y}]$  in the space  $L$ , defined by an arbitrary Hermitian form  $\varphi(\mathbf{x}, \mathbf{y})$ ,

such that the associated quadratic form  $\varphi(\mathbf{x}, \mathbf{x})$  is positive definite:  $\varphi(\mathbf{x}, \mathbf{x}) > 0$  for every  $\mathbf{x} \neq \mathbf{0}$ . Let us now set

$$(\mathbf{x}, \mathbf{y}) = \sum_{g \in G} [\mathcal{A}_g(\mathbf{x}), \mathcal{A}_g(\mathbf{y})], \quad (14.10)$$

where the sum is taken over all elements  $g$  of the group  $G$ . We shall prove that  $(\mathbf{x}, \mathbf{y})$  is also a scalar product and that with respect to it, all transformations  $\mathcal{A}_g$  are unitary.

The required properties of a scalar product for  $(\mathbf{x}, \mathbf{y})$  derive from the analogous properties of  $[\mathbf{x}, \mathbf{y}]$  and from the fact that  $\mathcal{A}_g$  is a linear transformation:

1.  $(\mathbf{y}, \mathbf{x}) = \sum_{g \in G} [\mathcal{A}_g(\mathbf{y}), \mathcal{A}_g(\mathbf{x})] = \sum_{g \in G} \overline{[\mathcal{A}_g(\mathbf{x}), \mathcal{A}_g(\mathbf{y})]} = \overline{(\mathbf{x}, \mathbf{y})},$
2.  $(\lambda \mathbf{x}, \mathbf{y}) = \sum_{g \in G} [\mathcal{A}_g(\lambda \mathbf{x}), \mathcal{A}_g(\mathbf{y})] = \sum_{g \in G} \lambda [\mathcal{A}_g(\mathbf{x}), \mathcal{A}_g(\mathbf{y})] = \lambda (\mathbf{x}, \mathbf{y}),$
3.  $(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}) = \sum_{g \in G} [\mathcal{A}_g(\mathbf{x}_1 + \mathbf{x}_2), \mathcal{A}_g(\mathbf{y})]$   
 $= \sum_{g \in G} [\mathcal{A}_g(\mathbf{x}_1) + \mathcal{A}_g(\mathbf{x}_2), \mathcal{A}_g(\mathbf{y})] = (\mathbf{x}_1, \mathbf{y}) + (\mathbf{x}_2, \mathbf{y}),$
4.  $(\mathbf{x}, \mathbf{x}) = \sum_{g \in G} [\mathcal{A}_g(\mathbf{x}), \mathcal{A}_g(\mathbf{x})] > 0, \quad \text{if } \mathbf{x} \neq \mathbf{0}.$

For the proof of the last property, it is necessary to observe that in this sum, all terms  $[\mathcal{A}_g(\mathbf{x}), \mathcal{A}_g(\mathbf{x})]$  are positive. This follows from the analogous property of the scalar product  $[\mathbf{x}, \mathbf{y}]$ , that is, from the fact that  $[\mathbf{x}, \mathbf{x}] > 0$  for all  $\mathbf{x} \neq \mathbf{0}$ . Since the linear transformation  $\mathcal{A}_g : \mathbb{L} \rightarrow \mathbb{L}$  is nonsingular, it takes every nonnull vector  $\mathbf{x}$  to a nonnull vector  $\mathcal{A}_g(\mathbf{x})$ .

Let us now verify that with respect to the scalar product  $(\mathbf{x}, \mathbf{y})$ , every transformation  $\mathcal{A}_h$ ,  $h \in G$ , is unitary. In view of (14.10), we have

$$\begin{aligned} (\mathcal{A}_h(\mathbf{x}), \mathcal{A}_h(\mathbf{y})) &= \sum_{g \in G} [\mathcal{A}_g(\mathcal{A}_h(\mathbf{x})), \mathcal{A}_g(\mathcal{A}_h(\mathbf{y}))] \\ &= \sum_{g \in G} [\mathcal{A}_g \mathcal{A}_h(\mathbf{x}), \mathcal{A}_g \mathcal{A}_h(\mathbf{y})]. \end{aligned} \quad (14.11)$$

Let us set  $gh = u$ . In view of property (14.1), we have  $\mathcal{A}_g \mathcal{A}_h = \mathcal{A}_{gh} = \mathcal{A}_u$ . Therefore, we may rewrite equality (14.11) in the form

$$(\mathcal{A}_h(\mathbf{x}), \mathcal{A}_h(\mathbf{y})) = \sum_{u=gh} [\mathcal{A}_u(\mathbf{x}), \mathcal{A}_u(\mathbf{y})]. \quad (14.12)$$

Let us now observe that as  $g$  runs through all elements of the group  $G$  while  $h$  is fixed, the element  $u = gh$  also runs through all elements of the group  $G$ . This follows from the fact that for every element  $u \in G$ , the element  $g = uh^{-1}$  satisfies the relationship  $gh = u$ , and that for distinct  $g_1$  and  $g_2$ , we thereby obtain distinct elements  $u_1$  and  $u_2$ .

Thus in equality (14.12), the element  $u$  runs through the entire group  $G$ , and we can rewrite this equality in the form

$$(\mathcal{A}_h(\mathbf{x}), \mathcal{A}_h(\mathbf{y})) = \sum_{g \in G} [\mathcal{A}_g(\mathbf{x}), \mathcal{A}_g(\mathbf{y})],$$

whence in view of definition (14.10), it follows that  $(\mathcal{A}_h(\mathbf{x}), \mathcal{A}_h(\mathbf{y})) = (\mathbf{x}, \mathbf{y})$ , that is, the transformation  $\mathcal{A}_h$  is unitary with respect to the scalar product  $(\mathbf{x}, \mathbf{y})$ .  $\square$

**Corollary 14.19** *If the space  $L$  of a representation of a finite group contains an invariant subspace  $L'$ , then it contains another invariant subspace  $L''$  such that  $L = L' \oplus L''$ .*

This follows directly from Lemma 14.17 and from Theorem 14.18.

**Corollary 14.20** *Every representation of a finite group is a direct sum of irreducible representations.*

*Proof* If the space  $L$  of our representation  $\mathcal{A}_g$  does not have an invariant subspace different from  $(0)$  and all of  $L$ , then this representation itself is irreducible, and our assertion is true (although trivially so). But if the space  $L$  has an invariant subspace  $L'$ , then by Corollary 14.19, there exists an invariant subspace  $L''$  such that  $L = L' \oplus L''$ .

Let us apply the same argument to each of the spaces  $L'$  and  $L''$ . Continuing this process, we will eventually come to a halt, since the dimensions of the obtained subspaces are continually decreasing. As a result, we arrive at such a decomposition (14.8) with some number  $r \geq 2$  such that the invariant subspaces  $L_i$  contain no invariant subspaces other than  $(0)$  and all of  $L_i$ . This means precisely that the representations  $\mathcal{A}_g^{(1)}, \dots, \mathcal{A}_g^{(r)}$  induced in the subspaces  $L_1, \dots, L_r$  by our representation  $\mathcal{A}_g$  are irreducible, and the representation  $\mathcal{A}_g$  decomposes as a direct sum  $\mathcal{A}_g^{(1)}, \dots, \mathcal{A}_g^{(r)}$ .  $\square$

**Theorem 14.21** *If a representation  $\mathcal{A}_g$  decomposes into a direct sum of irreducible representations  $\mathcal{A}_g^{(1)}, \dots, \mathcal{A}_g^{(r)}$ , then every irreducible representation  $\mathcal{B}_g$  contained in  $\mathcal{A}_g$  is equivalent to one of the  $\mathcal{A}_g^{(i)}$ .*

*Proof* Let  $L = L_1 \oplus \dots \oplus L_r$  be a decomposition of the space  $L$  of the representation  $\mathcal{A}_g$  into a direct sum of invariant subspaces such that  $\mathcal{A}_g$  induces in  $L_i$  the representation  $\mathcal{A}_g^{(i)}$ , and let  $M$  be the invariant subspace  $L$  in which  $\mathcal{A}_g$  induces the representation  $\mathcal{B}_g$ .

Then in particular, for every vector  $\mathbf{x} \in M$ , we have the decomposition

$$\mathbf{x} = \mathbf{x}_1 + \dots + \mathbf{x}_r, \quad \mathbf{x}_i \in L_i. \quad (14.13)$$

It determines a linear transformation  $\mathcal{P}_i : M \rightarrow L_i$  that is the projection of the subspace  $M$  onto  $L_i$  parallel to  $L_1 \oplus \dots \oplus L_{i-1} \oplus L_{i+1} \oplus \dots \oplus L_r$ ; see Example 3.51 on

p. 103. In other words, the transformations  $\mathcal{P}_i : M \rightarrow L_i$  are defined by the conditions

$$\mathcal{P}_i(\mathbf{x}) = \mathbf{x}_i, \quad i = 1, \dots, r. \quad (14.14)$$

The proof of the theorem is based on the relationships

$$\mathcal{A}_g \mathcal{P}_i(\mathbf{x}) = \mathcal{P}_i \mathcal{A}_g(\mathbf{x}), \quad i = 1, \dots, r, \quad (14.15)$$

which are valid for every vector  $\mathbf{x} \in M$ . For the proof of relationships (14.15), let us apply the transformation  $\mathcal{A}_g$  to both sides of equality (14.13). We then obtain

$$\mathcal{A}_g(\mathbf{x}) = \mathcal{A}_g(\mathbf{x}_1) + \dots + \mathcal{A}_g(\mathbf{x}_r). \quad (14.16)$$

Since  $\mathcal{A}_g(\mathbf{x}) \in M$  and  $\mathcal{A}_g(\mathbf{x}_i) \in L_i$ ,  $i = 1, \dots, r$ , it follows that relationship (14.16) is decomposition (14.13) for the vector  $\mathcal{A}_g(\mathbf{x})$ , whence follows equality (14.15).

From the irreducibility of the representations  $\mathcal{A}_g^{(1)}, \dots, \mathcal{A}_g^{(r)}$  and  $\mathcal{B}_g$ , it follows that the projection  $\mathcal{P}_i$  defined by formula (14.14) is either identically zero or an isomorphism of the spaces  $M$  and  $L_i$ . Indeed, let the vector  $\mathbf{x} \in M$  be contained in the kernel of the transformation  $\mathcal{P}_i$ , that is,  $\mathcal{P}_i(\mathbf{x}) = \mathbf{0}$ . Then clearly,  $\mathcal{A}_g \mathcal{P}_i(\mathbf{x}) = \mathbf{0}$ , and in view of relationship (14.15), we obtain that  $\mathcal{P}_i \mathcal{A}_g(\mathbf{x}) = \mathbf{0}$ , that is, the vector  $\mathcal{A}_g(\mathbf{x})$  is also contained in the kernel of  $\mathcal{P}_i$ . From the irreducibility of the representations  $\mathcal{A}_g^{(i)}$ , it now follows that the kernel either is equal to  $(\mathbf{0})$  or coincides with the entire space  $M$  (in the latter case, the projection  $\mathcal{P}_i$  will obviously be the null transformation). In exactly the same way, from equality (14.15), it follows that the image of the transformation  $\mathcal{P}_i$  either equals  $(\mathbf{0})$  or coincides with the subspace  $L_i$ .

However, there is certainly at least one such index  $i$  among the numbers  $1, \dots, r$  for which the transformation  $\mathcal{P}_i$  is not identically zero. For this, we must take an arbitrary nonnull vector  $\mathbf{x} \in M$  one of whose components  $\mathbf{x}_i$  in the decomposition (14.13) is not equal to zero, and therefore,  $\mathcal{P}_i(\mathbf{x}) \neq \mathbf{0}$ . Taking into account the previous arguments, this shows that the corresponding transformation  $\mathcal{P}_i$  is an isomorphism of the vector spaces  $M$  and  $L_i$ , and relationship (14.15) shows the equivalence of the corresponding representations  $\mathcal{B}_g$  and  $\mathcal{A}_g^{(i)}$ .  $\square$

**Corollary 14.22** *In a given representation are contained only finitely many distinct—in the sense of equivalence—irreducible representations.*

Indeed, all irreducible representations contained in the given one are equivalent to one of those encountered in an arbitrary decomposition of this representation as a direct sum of irreducible representations.

**Remark 14.23** From Theorem 14.21 there follows a certain property of *uniqueness* of the decompositions of a representation into irreducible representations. Namely, however we decompose a representation, we shall encounter in the decomposition the same (up to equivalence) irreducible representations. Indeed, let us select a certain decomposition of our representation into irreducible representations. An irreducible representation encountered in any other decomposition appears in our representation, which means that by Theorem 14.21, it is equivalent to one of the terms

of the chosen decomposition. A stronger property of uniqueness consists in the fact that if in one decomposition there appear  $k$  terms equivalent to a given irreducible representation, then the same number of such terms will appear as well in every other decomposition. We shall not require this assertion in the sequel, and we shall therefore not prove it.

### 14.3 Irreducible Representations

In this section, we shall prove that a finite group has only a finite number of distinct (up to equivalence) irreducible representations. We shall accomplish this as follows: We shall construct one particularly important representation called a *regular representation*, for which we then shall prove that every irreducible representation is contained within it. The finiteness of the number of such representations will then result from Corollary 14.22. The space of a regular representation consists of *all possible functions on the group*. This is a special case of the general notion of the space of functions on an arbitrary set (see Example 3.36, p. 94).

For an arbitrary finite group  $G$ , let us consider the vector space  $M(G)$  of functions on this group. Since the group  $G$  is finite, the space  $M(G)$  has finite dimension:  $\dim M(G) = |G|$ .

**Definition 14.24** The *regular* representation of a group  $G$  is the representation  $\mathcal{R}_g$  whose representation space is the space  $M(G)$  of functions on the group  $G$ , and in which the element  $g \in G$  is associated with the linear transformation  $\mathcal{R}_g$  that takes the function  $f(h) \in M(G)$  to the function  $\varphi(h) = f(hg)$ :

$$(\mathcal{R}_g(f))(h) = f(hg). \quad (14.17)$$

Formula (14.17) means that the result of applying the linear transformation  $\mathcal{R}_g$  to the function  $f$  is a “translated” function  $f$ , in the sense that the value  $\mathcal{R}_g(f)$  on the element  $h \in G$  is equal to  $f(hg)$ . We shall omit the obvious verification of the fact that the transformation of the space  $M(G)$  thus obtained is linear. Let us verify that  $\mathcal{R}_g$  is a representation, that is, that it satisfies the requirements (14.1).

Let us set  $\mathcal{R}_{g_1 g_2}(f) = \varphi$ . By formula (14.17), we have

$$\varphi(h) = f(hg_1 g_2).$$

Let  $\mathcal{R}_{g_2}(f) = \psi$ . Then

$$\psi(u) = f(ug_2).$$

Finally, if  $\mathcal{R}_{g_1} \mathcal{R}_{g_2}(f) = \varphi_1$ , then  $\varphi_1 = \mathcal{R}_{g_1}(\psi)$  and  $\varphi_1(u) = \psi(ug_1)$ . Substituting  $u = hg_1$  into the previous formula, we obtain that  $\varphi_1(u) = \psi(ug_1) = f(ug_1 g_2)$  for every element  $u \in G$ . This means that  $\varphi = \varphi_1$  and  $\mathcal{R}_{g_1 g_2} = \mathcal{R}_{g_1} \mathcal{R}_{g_2}$ .

*Example 14.25* Let  $G$  be a group of order two, consisting of elements  $e$  and  $g$ , where  $g^2 = e$ . A particular instance of this group is  $S_2$ , the symmetric group of

degree 2. The space  $M(G)$  is two-dimensional, and every function  $f \in M(G)$  is defined by two numbers,  $\alpha = f(e)$  and  $\beta = f(g)$ , that is, it can be identified with the vector  $(\alpha, \beta)$ . As with any representation,  $\mathcal{R}_e$  is the identity transformation. Let us determine what  $\mathcal{R}_g$  is. By formula (14.17), we have

$$(\mathcal{R}_g(f))(e) = f(g) = \beta, \quad (\mathcal{R}_g(f))(g) = f(g^2) = f(e) = \alpha.$$

This means that the linear transformation  $\mathcal{R}_g$  takes the vector  $(\alpha, \beta)$  to the vector  $(\beta, \alpha)$ , that is, it represents a reflection with respect to the line  $\alpha = \beta$ .

**Theorem 14.26** *Every irreducible representation of a finite group  $G$  is contained in its regular representation  $\mathcal{R}_g$ .*

*Proof* Let  $\mathcal{A}_g$  be an irreducible representation with space  $L$ . Let us denote by  $l$  an arbitrary nonnull linear function on the space  $L$  and let us associate with each vector  $\mathbf{x} \in L$  the function  $f(h) = l(\mathcal{A}_h(\mathbf{x})) \in M(G)$  obtained when the vector  $\mathbf{x}$  is fixed and the element  $h$  runs through all possible values of the group  $G$ . It is obvious that in this way, we obtain a linear transformation  $\mathcal{C} : L \rightarrow M'$  defined by the relationship

$$\mathcal{C}(\mathbf{x}) = l(\mathcal{A}_h(\mathbf{x})), \quad (14.18)$$

where  $M'$  is some subspace of the vector space  $M(G)$ . Here by construction,  $\mathcal{C}(L) = M'$ , that is,  $M'$  is the image of the transformation  $\mathcal{C}$ .

We shall prove the following properties:

- (1) For all elements  $g \in G$  and vectors  $\mathbf{x} \in L$ , we have the relationship

$$(\mathcal{C}\mathcal{A}_g)(\mathbf{x}) = (\mathcal{R}_g\mathcal{C})(\mathbf{x}). \quad (14.19)$$

- (2) The subspace  $M'$  is invariant with respect to the representation  $\mathcal{R}_g$ .  
 (3) The transformation  $\mathcal{C}$  is an isomorphism of the spaces  $L$  and  $M'$ .

Comparing formulas (14.19) and (14.7), taking into account the remaining two properties, we conclude that the irreducible representation  $\mathcal{A}_g$  is equivalent to the representation induced by the regular representation  $\mathcal{R}_g$  in the invariant subspace  $M' \subset M(G)$ . By virtue of the definitions given above, this means that  $\mathcal{A}_g$  is contained in  $\mathcal{R}_g$ , as asserted in the statement of the theorem.

*Proof of property (1).* Let us set  $\mathcal{C}(\mathbf{x}) = f \in M(G)$ . Then by definition,  $f(h) = l(\mathcal{A}_h(\mathbf{x}))$  for every element  $h \in G$ . Applying formula (14.17), we obtain the relationship

$$(\mathcal{R}_g\mathcal{C})(\mathbf{x}) = \mathcal{R}_g(f) = \varphi, \quad (14.20)$$

where  $\varphi$  is the function on the group  $G$  defined by the relationship  $\varphi(h) = l(\mathcal{A}_{hg}(\mathbf{x}))$ .

On the other hand, substituting the vector  $\mathcal{A}_g(\mathbf{x})$  for  $\mathbf{x}$  in formula (14.18), we obtain the equality

$$\mathcal{C}(\mathcal{A}_g(\mathbf{x})) = (\mathcal{C}\mathcal{A}_g)(\mathbf{x}) = \varphi_1(h), \quad (14.21)$$



where the function  $\varphi_1(h)$  is defined by the relationship

$$\varphi_1(h) = l(\mathcal{A}_h \mathcal{A}_g(\mathbf{x})) = l(\mathcal{A}_{hg}(\mathbf{x})),$$

and clearly, it coincides with  $\varphi(h)$ . Taking into account that  $\varphi(h) = \varphi_1(h)$ , we see that equalities (14.20) and (14.21) yield that  $(\mathcal{C} \mathcal{A}_g)(\mathbf{x}) = (\mathcal{R}_g \mathcal{C})(\mathbf{x})$ .

*Proof of property (2).* We must prove that for every element  $g \in G$ , the image of the linear transformation  $\mathcal{R}_g(M')$  is contained in  $M'$ . Let  $f \in M'$ , that is, by the definition of the image,  $f = \mathcal{C}(\mathbf{x})$  for some  $\mathbf{x} \in L$ . Then taking into account formula (14.19) proved above, we have the equality

$$\mathcal{R}_g(f) = (\mathcal{R}_g \mathcal{C})(\mathbf{x}) = (\mathcal{C} \mathcal{A}_g)(\mathbf{x}) = \mathcal{C}(\mathbf{y}),$$

where the vector  $\mathbf{y} = \mathcal{A}_g(\mathbf{x})$  is in  $L$ , and by our construction, this means that  $\mathcal{R}_g(f) \in M'$ . This proves the required inclusion  $\mathcal{R}_g(M') \subset M'$ .

*Proof of property (3).* Since by construction, the space  $M'$  is the image of the transformation  $\mathcal{C} : L \rightarrow M'$ , it remains only to show that the transformation  $\mathcal{C}$  is bijective, that is, that its kernel is equal to  $(\mathbf{0})$ . This means that we must prove that the equality  $\mathbf{x} = \mathbf{0}$  follows from the equality  $\mathcal{C}(\mathbf{x}) = \mathbf{0}'$  (where  $\mathbf{0}'$  denotes the function identically equal to zero on the group  $G$ ). Let us denote the kernel of the transformation  $\mathcal{C}$  by  $L'$ . As we know, it is a subspace of  $L$ . Let us show that  $L'$  is invariant with respect to the representation  $\mathcal{A}_g$ .

Indeed, let us suppose that  $\mathcal{C}(\mathbf{x}) = \mathbf{0}'$  for some vector  $\mathbf{x} \in L$ , and let us set  $\mathbf{y} = \mathcal{A}_g(\mathbf{x})$ . On applying the transformation  $\mathcal{C}$  to the vector  $\mathbf{y}$ , taking into account formula (14.19), we obtain

$$\mathcal{C}(\mathbf{y}) = (\mathcal{C} \mathcal{A}_g)(\mathbf{x}) = (\mathcal{R}_g \mathcal{C})(\mathbf{x}) = \mathcal{R}_g(\mathcal{C}(\mathbf{x})) = \mathcal{R}_g(\mathbf{0}') = \mathbf{0}'.$$

But from the irreducibility of the representation  $\mathcal{A}_g$ , it now follows that either  $L' = L$  or  $L' = (\mathbf{0})$ . The former would mean that  $l(\mathcal{A}_h(\mathbf{x})) = 0$  for all  $h \in G$  and  $\mathbf{x} \in L$ . But then even for  $h = e$ , we would have the equality  $l(\mathcal{A}_e(\mathbf{x})) = l(\mathcal{E}(\mathbf{x})) = l(\mathbf{x}) = 0$  for all  $\mathbf{x} \in L$ , which is impossible, since in the definition of the transformation  $\mathcal{C}$ , the function  $l$  was chosen to be not identically zero. This means that the subspace  $L'$  is equal to  $(\mathbf{0})$ , which is what was to be proved.  $\square$

**Corollary 14.27** *A finite group has only a finite number of distinct (up to equivalence) irreducible representations.*

*Example 14.28* Let  $\mathcal{A}_g$  be the one-dimensional identity representation of the group  $G$ . Then the space  $L$  is one-dimensional. Let  $\mathbf{e}$  be a basis of  $L$ . Let us define the function  $l$  by the condition  $l(\alpha \mathbf{e}) = \alpha$ . Formula (14.18) gives for the vector  $\mathbf{x} = \alpha \mathbf{e}$ , the value

$$\mathcal{C}(\alpha \mathbf{e}) = f, \quad \text{where } f(h) = l(\mathcal{A}_h(\alpha \mathbf{e})) = l(\alpha \mathbf{e}) = \alpha.$$

Thus to the vector  $\alpha \mathbf{e}$  is associated the function  $f$ , which takes for all  $h \in G$  the same value  $\alpha$ . Obviously, such constant functions indeed form an invariant subspace with respect to the regular representation, and the representation induced in it is the identity, as asserted by Theorem 14.26.

## 14.4 Representations of Abelian Groups

Let us first of all recall that we are assuming throughout that the space  $L$  of a representation is complex.

**Theorem 14.29** *An irreducible representation of an abelian group is one-dimensional.*

*Proof* Let  $g$  be a fixed element of the group  $G$ . Its associated linear transformation  $\mathcal{A}_g : L \rightarrow L$  has at least one eigenvalue  $\lambda$ . Let  $M \subset L$  be the eigensubspace corresponding to the eigenvalue  $\lambda$ , that is, the collection of all vectors  $x \in L$  such that

$$\mathcal{A}_g(x) = \lambda x. \quad (14.22)$$

By construction,  $M \neq (0)$ . We shall now prove that  $M$  is an invariant subspace of our representation. It will then follow from the irreducibility of the representation that  $M = L$ , and then equality (14.22) will hold for every vector  $x \in L$ . In other words,  $\mathcal{A}_g = \lambda E$ , and the matrix of the transformation  $\mathcal{A}_g$  is equal to  $\lambda E$ . A matrix of this type is called a *scalar matrix*. This reasoning holds for every  $g \in G$ ; we have only to note that the eigenvalue  $\lambda$  in formula (14.22) depends on the element  $g$ , and the remainder of the argument does not depend on it. Thus we may conclude that the matrices of all transformations  $\mathcal{A}_g$  are scalar matrices, and if  $\dim L > 1$ , then every subspace of the space  $L$  is invariant. Consequently, if a representation is irreducible, it is one-dimensional.

It remains to prove the invariance of the subspace  $M$ . It is here that we shall specifically use the commutativity of the group  $G$ . Let  $x \in M$ ,  $h \in G$ . We shall prove that  $\mathcal{A}_h(x) \in M$ . Indeed, if  $\mathcal{A}_h(x) = y$ , then

$$\begin{aligned} \mathcal{A}_g(y) &= \mathcal{A}_g(\mathcal{A}_h(x)) = \mathcal{A}_{gh}(x) = \mathcal{A}_{hg}(x) = \mathcal{A}_h(\mathcal{A}_g(x)) = \mathcal{A}_h(\lambda x) \\ &= \lambda \mathcal{A}_h(x) = \lambda y, \end{aligned}$$

that is, the vector  $y$  belongs to  $M$ . □

In view of Theorem 14.29, every irreducible representation of an abelian group can be represented in the form  $\mathcal{A}_g = \chi(g)$ , where  $\chi(g)$  is a number. Condition (14.1) can then be written in the following form:

$$\chi(g_1 g_2) = \chi(g_1) \chi(g_2). \quad (14.23)$$

**Definition 14.30** A function  $\chi(g)$  on an abelian group  $G$  taking complex values and satisfying relationship (14.23) is called a *character*.

By Theorem 14.29, every irreducible representation of a finite abelian group is a character  $\chi(g)$ . On the other hand, it follows from Theorem 14.26 that this representation is contained in the regular representation. In other words, in the space  $M(G)$  of functions on the group  $G$ , there exists an invariant subspace  $M'$  in which

the regular representation induces a representation equivalent to ours. Since our representation is one-dimensional, the subspace  $M'$  is also one-dimensional. Let some function  $f \in M(G)$  be a basis in  $M'$ . Then since the representation induced by the regular representation in  $M'$  has matrix  $\chi(g)$ , and  $\mathcal{R}_g(f)(h) = f(hg)$ , we must have the relationship

$$f(hg) = \chi(g)f(h).$$

Let us set  $h = e$  in this equality and let us also set  $f(e) = \alpha$ . We obtain that  $f(g) = \alpha\chi(g)$ , that is, we may take as a basis of the subspace  $M'$  the character  $\chi$  itself (indeed, it is a function on  $G$ , and this means that  $\chi \in M(G)$ ). As we have seen, we then have  $M(G) = M' \oplus M''$ , where  $M''$  is also an invariant subspace. Applying analogous arguments to  $M''$  and to all invariant subspaces of dimension greater than 1 that we obtain along the way, we finally arrive at a decomposition of the subspace  $M(G)$  as a direct sum of one-dimensional invariant subspaces. We have thereby proved the following result.

**Theorem 14.31** *The space  $M(G)$  of functions on a finite abelian group  $G$  can be decomposed as a direct sum of one-dimensional subspaces that are invariant with respect to the regular representation. In each such subspace, one can take as a basis vector some character  $\chi(g)$ . Then the matrix of the representation that is induced in this subspace coincides with this same character  $\chi(g)$ .*

It is obvious that we thereby establish a bijective relationship between the characters of the group  $G$  and one-dimensional invariant subspaces of the space  $M(G)$  of functions on this group. Indeed, two distinct characters  $\chi_1$  and  $\chi_2$  cannot be basis vectors of one and the same representation: that would mean that

$$\chi_1(g) = \alpha\chi_2(g) \quad \text{for all } g \in G.$$

Setting here  $g = e$ , we obtain  $\alpha = 1$ , since  $\chi_1$  and  $\chi_2$  are homomorphisms of the group  $G$  into  $\mathbb{C}$ , and therefore,  $\chi_1(e) = \chi_2(e) = 1$ .

Since by Corollary 14.19, a regular representation can be decomposed into a direct sum of irreducible representations, we obtain the following results for every finite abelian group  $G$ .

**Corollary 14.32** *The characters form a basis of the space  $M(G)$  of functions on the group  $G$ .*

This assertion can be reformulated as follows.

**Corollary 14.33** *The number of distinct characters of a group  $G$  is equal to its order.*

This follows from Corollary 14.32 and the fact that the dimension of the space  $M(G)$  is equal to the order of the group  $G$ .

**Corollary 14.34** *Every function on the group  $G$  is a linear combination of characters.*

*Example 14.35* Let  $G = \{g\}$  be a cyclic group of finite order  $n$ ,  $g^n = e$ . Let us denote by  $\xi_0, \dots, \xi_{n-1}$  the distinct  $n$ th roots of 1, and let us set

$$\chi_i(g^k) = \xi_i^k, \quad k = 0, 1, \dots, n-1.$$

It is easily verified that  $\chi_i$  is a character of the group  $G$  and that the characters  $\chi_i$  corresponding to  $\xi_i$ , the distinct  $n$ th roots of 1, are themselves distinct. Since their number is equal to  $|G|$ , they must be all the characters of the group  $G$ . By Corollary 14.32, they form a basis of the space  $M(G)$ . In other words, in an  $n$ -dimensional space, the vectors  $1, \xi_i, \dots, \xi_i^{n-1}$  corresponding to the  $n$ th roots of 1 form a basis. This can also be verified directly by calculating the determinant consisting of the coordinates of these vectors as a Vandermonde determinant (p. 41).

*Example 14.36* Let us denote by  $S$  the group of rotations of the circle in the plane. The elements of the group  $S$  correspond to points of the circle: if we associate with a real number  $\varphi$  the point of the circle with argument  $\varphi$ , then with any one point of the circle will be associated numbers that differ from one another by an integer multiple of  $2\pi$ . Therefore, this group  $S$  is frequently called the *circle group*.

After choosing a certain integer  $m$ , let us associate with the point  $t$  of the circle  $S$  having argument  $\varphi$  the number  $\cos m\varphi + i \sin m\varphi$ , where  $i$  is the imaginary unit. It is obvious that adding an integer multiple of  $2\pi$  to  $\varphi$  does not change this number, which means that it is uniquely defined by the point  $t \in S$ . Let us set

$$\chi_m(t) = \cos m\varphi + i \sin m\varphi, \quad m = 0, \pm 1, \pm 2, \dots \quad (14.24)$$

It is not difficult to verify that the function  $\chi_m(t)$  is a character of the group  $S$ . For an infinite group such as  $S$ , it is natural to introduce into the definition of a character in addition to the requirement (14.23), the requirement that the function  $\chi_m(t)$  be continuous. The reason for such a requirement for the group  $S$  is as follows: it is necessary that the real and complex parts of the functions  $\chi_m(t)$  be continuous functions.

It is possible to prove that the characters  $\chi_m(t)$  defined by formula (14.24) are continuous and that they comprise all the continuous characters of the circle. This explains to a large degree the role of the trigonometric functions  $\cos m\varphi$  and  $\sin m\varphi$  in mathematics: they are the real and imaginary parts of the continuous characters of the circle.

Corollary 14.34 asserts that every function on a finite abelian group can be represented as a linear combination of characters. In the case of an infinite group such as  $S$ , some analytic restrictions, which we shall not specify here, are naturally imposed on such a function. We shall only mention the significance of functions on the group  $S$ . Such a function  $f(t)$  can be represented as a function  $F(\varphi)$  of the argument  $\varphi$  of the point  $t \in S$ . It must not, however, depend on the choice of the argument  $\varphi$  of the point  $t$ , that is, it must not change on the addition to  $\varphi$  of an integer multiple of  $2\pi$ . In other words,  $F(\varphi)$  must be a periodic function with period  $2\pi$ .

The analogue of Corollary 14.34 for the group  $S$  asserts that such a function can be represented as a linear combination (in the given case, infinite) of functions  $\chi_m(\varphi)$ ,  $m = 0, \pm 1, \pm 2, \dots$ . In other words, this is a theorem about the fact that a periodic function (with certain analytic restrictions) can be decomposed into a Fourier series.

## Historical Note

Here we shall present a brief chronology of the appearance of the concepts discussed in this book. The development of mathematical ideas generally proceeds in such a way that some concepts gradually emerge from others. Therefore, it is generally impossible to fix accurately the appearance of some particular idea. We shall only point out the important milestones and, it goes without saying, shall do so only roughly. In particular, we shall limit our view to Western European mathematics.

The principal stimulus was, of course, the creation of analytic geometry by Fermat and Descartes in the seventeenth century. This made it possible to specify points (on the line, in the plane, and in three-dimensional space) using numbers (one, two, or three), to specify curves and surfaces by equations, and to classify them according to the algebraic nature of their equations. In this regard, linear transformations were used frequently, especially by Euler, in the eighteenth century.

Determinants (particularly as a symbolic apparatus for finding solutions of systems of  $n$  linear equations in  $n$  unknowns) were considered by Leibniz in the seventeenth century (even if only in a private letter) and in detail by Gabriel Cramer in the eighteenth. It is of interest that they were constructed on the basis of the rule of “general expansion” of the determinant, that is, on the basis of the most complex (among those that we considered in Chap. 2) way of defining them. This definition was discovered “empirically,” that is, conjectured on the basis of the formulas for the solution of systems of linear equations in two and three unknowns. The broadest use of determinants occurred in the nineteenth century, especially in the work of Cauchy and Jacobi.

The concept of “multidimensionality,” that is, the passage from one, two, and three coordinates to an arbitrary number, was stimulated by the development of mechanics, where one considered systems with an arbitrary number of degrees of freedom. The idea of extending geometric intuition and concepts to this case was developed systematically by Cayley and Grassmann in the nineteenth century. At the same time, it became clear that one must study quadrics in spaces of arbitrary dimension (Jacobi and Sylvester in the nineteenth century). In fact, this question had already been considered by Euler.

The study of concepts defined by a set of abstract axioms (groups, rings, algebras, fields) began as early as the nineteenth century in the work of Hamilton and Cayley, but it reached its full flowering in the twentieth century, chiefly in the schools of Emmy Noether and Emil Artin.

The concept of a projective space was first investigated by Desargues and Pascal in the seventeenth century, but systematic work in this direction began only in the nineteenth century, beginning with the work of Poncelet.

The axiomatic definition of vector spaces and Euclidean spaces as given in this book broke finally with the primacy of coordinates. It was first rigorously formulated almost simultaneously by Hermann Weyl and John von Neumann. Both came to this from work on questions in physics. Then two versions of quantum mechanics were created: the “wave mechanics” of Schrödinger and the “matrix mechanics” of Heisenberg. It was necessary to work out that in some sense, they were “one and the same.”

Both mathematicians developed an axiomatic theory of Euclidean spaces and vector spaces and showed that quantum-mechanical theories are connected with two isomorphic spaces. However, the difference between those theories and what we presented in this book lies in the fact that they worked with infinite-dimensional spaces. In any case, for finite-dimensional spaces, there appeared an invariant (that is, independent of the choice of coordinates) theory that by now has become universally accepted.

The introduction of the axiomatic approach in geometry was discussed in sufficient detail in Chap. 11, devoted to the hyperbolic geometry of Lobachevsky. Such studies began at the end of the nineteenth century, but their definitive influence in mathematics dates from the beginning of the twentieth century. The central figure here was Hilbert. For example, he contributed to the application of geometric intuition to many problems in analysis.

# References

We recall first those books that were in vogue when the lectures on which this book is based were given. Many of these books have been reprinted, and we have tried to provide information on the latest available version.<sup>1</sup>

1. I.M. Gelfand, *Lectures on Linear Algebra* (Dover, New York, 1989)
2. A.G. Kurosh, *Linear Equations from a Course of Higher Algebra* (Oregon State University Press, Corvallis, 1969)
3. F.R. Gantmacher, *The Theory of Matrices* (American Mathematical Society, Chelsea, 1959)
4. A.I. Malcev, *Foundations of Linear Algebra* (Freeman, New York, 1963)
5. P.R. Halmos, *Finite-Dimensional Vector Spaces* (Springer, New York, 1974)
6. G.E. Shilov, *Mathematical Analysis: A Special Course* (Pergamon, Elmsford, 1965)
7. O. Schreier, E. Sperner, *Introduction to Modern Algebra and Matrix Theory*, 2nd edn. (Dover, New York, 2011)
8. O. Schreier, E. Sperner, *Einführung in die analytische Geometrie und Algebra* (Teubner, Leipzig, 1931)

The book by Shilov is of particular interest for its large number of analytic applications. The following books could also be recommended. However, the conciseness of their presentation and abstract approach put them far beyond the capacity of the average student.

9. B.L. Van der Waerden, *Algebra* (Springer, New York, 2003)
10. N. Bourbaki, *Algebra I* (Springer, Berlin, 1998)
11. N. Bourbaki, *Algebra II* (Springer, Berlin, 2003)

Since the lectures on which this book is based were given, so many books on the subject have appeared that we give here only a small sample.

---

<sup>1</sup>*Translator's note:* Wherever possible, English-language versions have been given. Some of these were written originally in English, while others are translations from original Russian or German sources.



12. E.B. Vinberg, *A Course in Algebra* (American Mathematical Society, Providence, 2003)
13. A.I. Kostrikin, Yu.I. Manin, *Linear Algebra and Geometry* (CRC Press, Boca Raton, 1989)
14. A.I. Kostrikin, *Exercises in Algebra: A Collection of Exercises* (CRC Press, Boca Raton, 1996)
15. S. Lang, *Algebra* (Springer, 1992)
16. M.M. Postnikov, *Lectures in Geometry: Semester 2* (Mir, Moscow, 1982)
17. D.K. Faddeev, *Lectures on Algebra* (Lan, St. Petersburg, 2005) (in Russian)
18. R.A. Horn, C.R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge 1990)

With regard to applications to mechanics, see the book *The Theory of Matrices* by Gantmacher mentioned above as well as the following.

19. F.R. Gantmacher, *Oscillation Matrices and Kernels and Small Vibrations of Mechanical Systems* (American Mathematical Society, Providence, 2002)

Relationships with differential geometry, which we briefly touched on in this course, are described, for example, in the following.

20. A.S. Mishchenko, A.T. Fomenko, *A Course of Differential Geometry and Topology* (Mir, Moscow, 1988)

In presenting Lobachevsky's hyperbolic geometry, we have followed for the most part the following brochure.

21. B.N. Delone, *Elementary Proof of the Consistency of Hyperbolic Geometry* (State Technical Press, Moscow, 1956) (in Russian)

All the results concerning the foundations of geometry whose proofs we omitted are contained in the following books.

22. N.Yu. Netsvetayev, A.D. Alexandrov, *Geometry* (Nauka, Fizmatlit, Moscow, 1990)
23. N.V. Efimov, *Higher Geometry* (Mir, Moscow, 1980)

Facts about analytic geometry that were briefly mentioned in this course, such as the connection with the theory of quadrics, can be found in the following books.

24. P. Dandelin, *Mémoire sur l'hyperboloïde de révolution, et sur les hexagones de Pascal et de M. Brianchon. Nouveaux mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles*, T. III (1826), pp. 3–16
25. B.N. Delone, D.A. Raikov, *Analytic Geometry* (State Technical Press, Moscow-Leningrad, 1949) (in Russian)
26. P.S. Alexandrov, *Lectures in Analytic Geometry* (Nauka, Fizmatlit, Moscow, 1968) (in Russian)
27. D. Hilbert, S. Cohn-Vossen, *Geometry and the Imagination* (AMS, Chelsea, 1999)
28. A.P. Veselov, E.V. Troitsky, *Lectures in Analytic Geometry* (Lan, St. Petersburg, 2003) (in Russian)

Connections between the hyperbolic geometry of Lobachevsky and other branches of projective geometry are well described in the following book.

29. F. Klein, *Nicht-Euklidische Geometrie* (Göttingen, 1893). Reprinted by AMS, Chelsea, 2000

In connection with representation theory, the following book is to be recommended.

30. J.-P. Serre, *Linear Representations of Finite Groups* (Springer, Berlin, 1977)

# Index

## A

Affine ratio  
 of three points, 298  
 Affine subset (of a projective space), 323  
 Affinely equivalent subsets, 307  
 Algebra, 370  
 exterior, 372  
 graded, 372  
 Angle  
 between planes, 237  
 between two lines or a line and a plane, 235  
 between vectors, 215  
 Annihilator, 124  
 Associativity, xv, 63, 371, 467  
 Axioms of plane geometry, 445  
 parallel lines (in Euclidean and hyperbolic geometry), 448

## B

Ball, 222  
 Bases  
 oriented, 155  
 with the same orientation, 277  
 Basis  
 of a vector space, 89  
 dual, 123  
 orthonormal (in a Euclidean space), 218  
 orthonormal (in a pseudo-Euclidean space), 266, 268  
 orthonormal (with respect to a bilinear form), 401  
 of an algebra, 371  
 Blocks of a matrix, 65

## C

Canonical equations (of a quadric), 422  
 Canonical form (of a quadratic form), 201

## Center

of a flag, 301, 442  
 of a set, 419  
 Central symmetry (of an affine space), 419  
 Character, 511  
 of the circle (continuous), 513  
 Characteristic polynomial, 139  
 Circle (group of rotations), 513  
 Cofactor, 40, 379  
 Combination  
 linear, 87  
 Commutativity, 473, 484  
 Commuting matrices, 64  
 Compactness, 341  
 Complexification, 151  
 Composition  
 of linear transformations, 106  
 of mappings, xiv  
 Cone  
 in an affine space, 421, 429  
 light (isotropic), 269  
 Conic, 392, 430  
 Constant terms, 1  
 Convergence, xviii, 179, 339  
 Coordinates  
 of a point, 291  
 heterogeneous, 323  
 of a vector, 90  
 Plücker (of a space), 351  
 points  
 homogeneous, 320  
 Cramer's rule, 43  
 Curvature  
 Gaussian, 265  
 normal, 263  
 principal, 264  
 Cylinder, 303

**D**

- Deformation (continuous), xx, 158, 343
- Degree of a polynomial, 15, 127
- Delta function, 94, 359
- Determinant, 25, 29
  - explicit formula, 53
  - Gram, 217
  - of a linear transformation, 112
  - of a square matrix, 30
  - Vandermonde, 41
- Diagonal (of a matrix), 2, 178
- Differential, 131, 293
- Dimension
  - of a projective space, 320
  - of a representation, 497
  - of a vector space, 88
  - of an affine space, 291
  - of an algebra, 371
- Direct sum
  - of representations, 502
  - of subgroups, 475
  - of submodules, 489
  - of subspaces, 84
- Distance between points, 309
- Distributive property, 64, 107, 370
- Divisor
  - greatest common (gcd), 487
  - (of an element of a ring), 486
- Duality principle, 125, 392

**E**

- Echelon form (systems of linear equations), 13
- Eigensubspace, 138
- Eigenvalue, 137
- Eigenvector, 137
- Element
  - identity, 370
  - inverse (right, left), 467
  - negative, 474
  - prime (of a ring), 486
  - torsion (in a module), 488
  - unit (in a ring), 486
  - zero, 474
- Elementary row operations (on matrices), 7
- Elements
  - associates (in a ring), 486
  - homogeneous (in a graded algebra), 373
  - relatively prime (in a ring), 487
- Ellipse, 430
- Ellipsoid, 428
- Endomorphism, 102
- Equivalence relation, xii
- Equivalent representations, 501
- Euclidean algorithm (in a ring), 487

## Exterior power

- $m$ th exterior power (of a vector space), 360

**F**

- Fiber of a projection, 303
- Field, 485
  - of characteristic different from 2, 83, 196
- Flag, 101, 301, 441, 447
- Form, 127
  - bilinear, 192
    - antisymmetric, symmetric, 193
    - nonsingular, 195
  - Hermitian, 210
  - quadratic, 191
    - first, second (of a hypersurface), 262
    - positive, negative definite, 205
  - sesquilinear, 210
- Formula
  - Cauchy–Binet, 377
  - change of basis
    - for the matrix of a bilinear form, 195
  - change of coordinates of a vector, 109
  - Euler, 264
  - expansion of the determinant along a
    - column, 40
  - for a change of matrix of a linear
    - transformation, 111
- Frame of reference, 291
  - orthonormal, 310
- Free mobility (of an affine Euclidean space), 317

## Function, xiii

- antisymmetric, 46
- exponential of a matrix, 181
- linear, 2
- multilinear, 51, 358
- quadratic Hermitian, 211
- semilinear, 209
- sesquilinear, 210
- symmetric, 44

**G**

- Gaussian elimination, 6
- Geometry
  - absolute, 448
  - elliptic, 464
  - projective, 319
  - spherical, 462
- Grade (of a principal vector), 162
- Grassmannian, 356
- Group, 467
  - abelian, 473
  - alternating of degree  $n$ , 471
  - commutative, 473

Group (*cont.*)

- cyclic, 471
- symmetric of degree  $n$ , 469
- transformation, 468

**H**

Half-space, 99, 436

## Hexagon

- circumscribed about a conic, 393
- inscribed in a conic, 392

Homeomorphism, xviii

Homomorphism (of groups), 471

Horizon, 324

Hyperbola, 430

Hyperboloid of one sheet, 398

Hyperplane, 89, 294, 322, 435  
tangent, 261, 327, 386

Hypersurface, 386

**I**

## Identity

- Cauchy–Binet, 68
- Euler's, 130

## Image

- of a homomorphism, 472
- of a linear transformation, 115
- of a mapping, xiii
- of an arbitrary mapping, xiii

Incidence (points and lines), 319

Index of inertia, 205, 266

## Inner product

- of vectors, 213, 435

Interpolation, 15

Inversion, 49

Isometry, xxi

## Isomorphism

- of affine spaces, 303
- of Euclidean spaces, 223
- of groups, 472
- of vector spaces, 112

**J**

## Jordan

- block, 169
- normal form, 169

**K**

## Kernel

- of a homomorphism, 472
- of a linear transformation, 115

**L**

Law of inertia, 205

Length of a vector, 215

Limit (of a sequence), xviii, 339

## Linear

- combination, 57
- part (of an affine transformation), 301
- substitution of variables, 62

**M**

## Mapping, xiii

- dual, xv
- extension, xiii
- identity, xiii, 102
- perspective, 338

## Matrices

- commuting, 64
- equivalent, 203
- similar, 135

## Matrix, 2

- additive inverse, 60
- adjugate, 73
- antisymmetric, 54
- block, 65
- block-diagonal, 65, 137
- continuously deformable, 158
- diagonal, 74
- echelon form, 13
- Hermitian, 210
- identity, 34
- inverse, 72
- nonsingular, 37
- null, 60
- of a bilinear form, 192
- of a linear transformation, 105
- orthogonal, 225
- singular, 37
- square, 2
- symmetric, 54
- system of linear equations, 2
- transition, 109
- transpose, 53

Metric, xvii, 309

## Minor, 31

- associates, 69
- leading principal, 206

Möbius strip, 346

Module (over a ring), 485

finitely generated, 488

## Motion

- in the axioms of plane geometry, 445
- of a hyperbolic space, 437
- of an affine Euclidean space, 310

Multiplication table (in an algebra), 371

$m$ -vector, 360

decomposable, 367

**N**

Newton sum, 209

Null vector, 81

**O**

Operations

in a group, 474

in a ring, 484

in an algebra, 370

Operator, 102

first-order differential, 129

Order

of a group, 468

of an element of a group, 471

of an element of a module, 489

Orientation

of a Euclidean space, 230

of a pseudo-Euclidean space, 277

of a vector space, 155

Orthogonal complement, 198, 218, 503

Orthonormal system of vectors, 218

**P**

Pair of half-spaces, 300

Parabola, 430

Parallel subspaces (in an affine space), 295

Parallelepiped (spanned by vectors), 219

Path (in a metric space), xx

Path-connected component, xx

Permutation, 45, 469

even, 48

Plücker relations, 354

Point

at infinity, 319, 324

critical, 253

fixed, 305

lying between two other points, 298, 445, 450

of a projective space, 320

of an affine space, 289

of hyperbolic space, 434

singular

of a hypersurface, 387

of a projective algebraic variety, 327

Points

independent, 297, 331

Poles (of the light cone), 271

Polynomial, 15, 127, 293

annihilator, 146, 147

characteristic, 139

homogeneous, 127

in a linear transformation, 141

matrix, 69

minimal, 146

Preimage, xiii

Principal of duality, 326

Product

direct

of subgroups, 474

of a matrix by a number, 60

of elements

of a group, 467

of an algebra, 370

of matrices, 61

of sets, xvi

of vectors

exterior, 360, 368

Projection, 103, 302

orthogonal, 216, 219

Projective

cover, 325

line, 320

plane, 320

Projectivization, 320

**Q**

Quadric, 385, 414

nonsingular, 386, 429

Quadrics

affinely equivalent, 418

metrically equivalent, 425

**R**

Radical (of a bilinear form), 198

Rank

of a bilinear form, 195

of a linear transformation, 118

of a matrix, 55

Ratio

of four points (cross, anharmonic), 337

Rectilinear generatrices (of a hyperboloid), 398

Reflection (of a Euclidean space), 229

Representation, 497

identity, 499

induced, 501

infinite-dimensional, 499

irreducible, reducible, 502

regular, 508

unitarizable, 504

Representation space, 497

Representations

equivalent, 501

Restriction (of a mapping), xiii

Ring, 484

commutative, 484

Euclidean, 486

Rotation of a Euclidean space about an axis,  
229

## S

Segment, 299, 446

Semiaxes (of an ellipsoid), 254, 428

Set, *xi*

centrally symmetric (in an affine space),  
419

convex (in an affine space), 299

Sets

homeomorphic, *xviii*

Solution of a system of linear equations, 4

Space

affine, 289

affine Euclidean, 309

dual, 121

Euclidean, 213

hyperbolic, 434

metric, *xvii*

Minkowski, 86, 268

*m*-vectors, 360

of a representation, 497

of linear functions, 121

of vectors of an affine space, 291

projective, 320

dual, 325

pseudo-Euclidean, 268

second dual, 123

tangent, 261, 327, 386

vector, 81

Sphere, 222

Stereographic projection, 343

Subgroup, 468

cyclic, 471

maximal, 476

Submodule, 488

cyclic, 489

Subspace

cyclic, 162

degenerate (of a pseudo-Euclidean space),  
266

invariant

(with respect to a linear  
transformation), 135

(with respect to a representation), 501

isotropic, 395

linear span of vectors, 87

nondegenerate (of a pseudo-Euclidean  
space), 266

of a hyperbolic space, 435

of a projective space, 322

dual, 326

of a vector space, 83

of an affine space, 294

solutions of a system of equations, 84

Subspaces

directed pair, 101

Sum

of matrices, 61

of subspaces, 84

direct, 84

Superalgebra, 373

Sylvester's criterion, 206

System of linear equations, 1

associated, 11

consistent, 5

definite, indefinite, 5

equivalent, 7

homogeneous, 10

inconsistent, 5

(row) echelon form, 13

uniquely determined, 5

upper triangular form, 14

## T

Theorem

Bolzano–Weierstrass, 247

Brianchon's, 393

Cayley–Hamilton, 147

Courant–Fischer, 253

Euler's, 316

Helmholtz–Lie, 443

Laplace's, 379

Pascal's, 392

Rouché–Capelli, 56

Torus, 414

Transformation

affine, 301

linear, 306

proper, improper, 307

singular, nonsingular, 304

antisymmetric, symmetric, 203, 245

block-diagonalizable, 152

diagonalizable, 139

dual, 125

linear, 102

Lorentz, 276

nonsingular, singular, 135

null, 106

of a vector space into itself, 133

orthogonal, 224, 401

projective, 328

proper, improper, 276, 402

singular, nonsingular, 111

unitary, 255, 503

Translation (of an affine space), 292

Transpose

- of a matrix, 53
- Transposition, 45
- Triangle, 446
- Triangle inequality (Cauchy–Schwarz), 310
  - in hyperbolic geometry, 458
  - in spherical geometry, 463
- U**
- Universality (of the exterior product), 365
- Unknowns
  - free, 13
  - principal, 13
- V**
- Variety
  - Grassmann, 356
  - projective algebraic, 322
- dual, 327
- irreducible, 409
- Vector, 79, 81
  - principal, 161
- Vectors
  - decomposable, 361
  - eigen-, 137
  - lightlike (isotropic), 269
  - linearly dependent, 87
  - linearly independent, 87
  - orthogonal, 198, 217
  - spacelike, 268
  - timelike, 269
- Volume of a parallelepiped
  - oriented, 221
  - unoriented, 220